STIC Database Tracking Number: 119347

TO: Minh-Tam Davis
Location: rem/3a24/3c18
Art Unit: 1642
Tuesday, April 13, 2004

Case Serial Number: 09/763335

From: Barb O'Bryen
Location: Biotech-Chem Library
Remsen 1A69
Phone: 571-272-2518

barbara.obryen@uspto.gov

Search Notes

Tam,
Paul Schulwitz assisted me with the Score/Length search. There were no hits that met your criteria of length between 60-200 nucleotides, and score over length value of 70% or greater. Consequently, there is no print-out for that segment of the search.
Please let me know if you have questions.
Barb

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: April 9, 2004, 09:44:26 ; Search time 397 Seconds (without alignments)
8689.010 Million cell updates/sec

Title:	US-09-763-335-4
Perfect score:	812
Sequence:	1 cactgcgagtggggatggtgcc.........cattcgtgg-cagtaaaggaa 812
Scoring table:	IDENTITY NUC
	Gapop 10.0 , Gapext 1.0
Searched:	3373863 seqs, 212409041 residues
Total number of hits satisfying chosen parameters:	6747726

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Geneseq_29Jan04:*
2: geneseqn1980s:*
3: geneseqn1990s:*
4: geneseqn2000s:*
5: geneseqn2001s:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match Length	Match Length	DB	DB ID	Description
1	812	100.0	812	3	AAZ51528	Human cel
2	686	84.5	2058	8	ABQ77548	Human pho
3	433.2	53.3	1936	5	ABX71301	Human kid
4	430.2	53.0	827	4	AAK51800	Human pol
5	425.4	52.4	671	4	AAK52784	Human pol
6	187	23.0	261	3	AAC08673	Human sec
7	131.2	16.2	716	4	AAI58347	Human pol
8	131.2	16.2	716	8	ADB48315	Novel hum
9	131.2	16.2	1171	4	AAI60133	Human pol
10	128.4	15.8	2390	9	ADB62738	Human cDN
11	109.6	13.5	238	2	AAC25659	Human sec
12	88.6	10.9	333	2	AAX51967	Human sec
13	70	8.6	429	4	AAH99019	Human EST
14	52.2	6.4	1493	3	AAZ33334	Human sec
15	52.2	6.4	2469	3	AAZ79939	Human sec
16	43.2	5.3	2000	7	ADA71938	Rice gene
17	42.6	5.2	16724	6	ABL33091	Human imm
18	42.6	5.2	16724	6	ABL34537	Human met
19	42.6	5.2	16724	6	ABL70260	Chemical1
20	42.6	5.2	301	5	AAA44625	Human sec
o 21	42	5.2	498	3	AAC94550	Cat flea
c 22	41.8	5.1	10886	6	ABL34135	Human imm
23	41.4	5.1	40862	6	ABL34073	Human imm
24	41	5.0	41	6	ABQ77553	Human pho
25	41	5.0	17869	6	ABK39921	Human che
26	41	5.0	17869	6	ABL32105	Human imm
27	40.4	5.0	13511	6	ABL32280	Human imm
28	40.2	5.0	4466	4	AAK69983	Human pol
29	40.2	4.9	9268	6	ABL33600	Human imm
30	39.8	4.9	2932	6	ABN85526	Human bis
31	39.8	4.9	2932	7	ABZ10173	Haematopo
32	39.8	4.9	2932	7	ABZ10027	Haematopo
33	39.8	4.9	2932	9	ADB54077	Pretreate
34	39.8	4.9	2932	9	ADB54205	Pretreate
35	39.8	4.9	2932	9	ADB84141	Human lym
36	39.8	4.9	2932	9	ADE84065	Human lym
37	39.8	4.9	13038	6	ABL33275	Human imm
38	39.6	4.9	398	4	AAI81224	Human pol
39	39.6	4.9	5875	6	ABL32289	Human imm
40	39.6	4.9	11416	6	ABL32118	Human imm
41	39.6	4.9	11416	6	ABL70135	Chemical1
42	39.6	4.9	11416	6	AAS61063	Human gen
43	39.4	4.9	41	6	ABQ77554	Human pho
44	39.4	4.9	89328	6	ABL61995	Colon ade
45	39.2	4.8	6904	6	ABL32213	Human imm

ALIGNMENTS

RESULT 1
AA251528

```
ID   AA251528 standard; cDNA; 812 BP.
XX
AC   AA251528;
XX
DT   21-JUN-2000  (first entry)
XX
DE   Human cell-signalling protein-2 cDNA.
XX
KW   Human cell-signalling protein-2; CSIG-2; BRAINOT12 cDNA library;
KW   neoplastic disorder; neurological disorder; immunological disorder;
KW   vesicle trafficking disorder; smooth muscle disorder; asthma; emphysema;
KW   rheumatoid arthritis; HIV; human immunodeficiency virus; osteoporosis;
KW   multiple sclerosis; atherosclerosis; diabetes mellitus; psoriasis;
KW   irritable bowel syndrome; viral infection; bacterial infection;
KW   fungal infection; cytostatic; antiarthritic; antiasthmatic;
KW   immunosuppressive; antiarteriosclerotic; anti-HIV; antidiabetic;
KW   antiinflammatory; neuroprotective; antipsoriatic; antimicrobial; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   CDS             385..786
FT                   /*tag= a
FT   sig_peptide     385..459
FT                   /product= "Cell-signalling protein-2"
FT                   /*tag= b
FT   mat_peptide     460..783
FT                   /*tag= c
FT                   /product= "Mature cell-signalling protein-2"
XX
PN   WO200011169-A2.
XX
PD   02-MAR-2000.
XX
PF   19-AUG-1999;  99WO-US019072.
XX
PR   21-AUG-1998;  98US-00137578.
PR   21-AUG-1998;  98US-0150689P.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Tang YT, Corley NC, Patterson C, Guegler KJ, Baughn MR;
XX
DR   WPI; 2000-245562/21.
```

ABQ77553 Human pho
ABK39921 Human che
ABL32105 Human imm
ABL32280 Human imm
AAK69983 Human pol
ABL33600 Human imm
ABN85526 Human bis
ABZ10173 Haematopo
ABZ10027 Haematopo
ADB54077 Pretreate
ADB54205 Pretreate
ADB84141 Human lym
ADE84065 Human lym
ABL33275 Human imm
AAI81224 Human pol
ABL32289 Human imm
ABL32118 Human imm
ABL70135 Chemical1
AAS61063 Human gen
ABQ77554 Human pho
ABL61995 Colon ade
ABL32213 Human imm

DR   P-PSDB; AAY70400.
XX
PT   New human cell signaling proteins and polynucleotides useful for
PT   diagnosis, prevention and treatment of neoplastic, neurological,
PT   immunological, vesicle trafficking and smooth muscle disorders.
XX
PS   Claim 9; Page 61; 62pp; English.
XX
CC   The present cDNA sequence encodes human cell-signalling protein-2 (CSIG-
CC   2) involved in regulation of of cell proliferation, differentiation and
CC   gene transcription. The CSIG-2 cDNA was isolated from brain tissue cDNA
CC   library (BRAINOT12). The BRAINOT12 library was constructed using RNA
CC   isolated from brain tissue removed from the right frontal lobe of a 5-
CC   year old Caucasian male during a hemispherectomy. CSIG cDNA can be used
CC   in the diagnosis and treatment of diseases associated with expression of
CC   CSIG. These diseases include neoplastic, neurological, immunological,
CC   vesicle trafficking and smooth muscle disorders, including HIV,
CC   rheumatoid arthritis, asthma, artherosclerosis, diabetes mellitus,
CC   emphysema, irritable bowel syndrome, multiple sclerosis, osteoporosis,
CC   psoriasis and infections including viral, bacterial and fungal.This is
CC   also useful for generating hybridisation probes useful in mapping the
CC   naturally occurring genomic sequences
XX
SQ   Sequence 812 BP; 243 A; 162 C; 195 G; 212 T; 0 U; 0 Other;

Query Match   100.0%; Score 812; DB 3; Length 812;
Best Local Similarity   100.0%; Pred. No. 2.6e-230;
Matches 812; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 CACTGGAGTGGGGATGGTCCATCGGGCGAACATATAAACTGATTCTCATCAGGAAAATGCACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CACTGGAGTGGGGATGGTCCATCGGGCGAACATATAAACTGATTCTCATCAGGAAAATGCACA 60

Qy   61 TTATCTCCCATCACTTCAAAGGTCTTCGTCAGGCAGGTGACCGCCAGGATGATTTTAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TTATCTCCCATCACTTCAAAGGTCTTCGTCAGGCAGGTGACCGCCAGGATGATTTTAA 120

Qy   121 AGGTGAAAATGACAAGGTTTCCACCCCTCAAACCTTGGCTCCTTTTCTGACAATACAGTC 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGGTGAAAATGACAAGGTTTCCACCCCTCAAACCTTGGCTCCTTTTCTGACAATACAGTC 180

Qy   181 TGAATGAACCCGATGTCTTTTTTTTTTACTGTGGAAATAGGATCGGAAGAGAGTAACATTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TGAATGAACCCGATGTCTTTTTTTTTTACTGTGGAAATAGGATCGGAAGAGAGTAACATTT 240

Qy   241 TTTTTTTAATCCTGATAAAGAAGATGTTGGGAAGCTCTTTGAAAAAAATTTTAAATTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TTTTTTTAATCCTGATAAAGAAGATGTTGGGAAGCTCTTTGAAAAAAATTTTAAATTG 300

Qy   301 TGGCACAGATGGATTTTAGATCTTTCCAATGAACACTAATAGAGTAGTACTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TGGCACAGATGGATTTTAGATCTTTCCAATGAACACTAATAGAGTAGTACTCT 360

Qy   361 GCTCTTGGCTGGATTTTTCAGGAATGGCTCTCGATGTCCTGGGTCCTGTAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GCTCTTGGCTGGATTTTTCAGGAATGGCTCTCGATGTCCTGGGTCCTGTAT 420

Qy   421 TTGTGGATAAGTGCTTGTGCAAATGCTACTCTGCCAGCACACTTTCCAG 480
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TTGTGGATAAGTGCTTGTGCAAATGCTACTCTGCCAGCACACTTTCCAG 480

Qy   481 CAGCATCACCTCACAGAACCAGAAGAGGAGGAGGCGTCACAAAGTAACACCGATGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CAGCATCACCTCACAGAACCAGAAGAGGAGGAGGCGTCACAAAGTAACACCGATGT 540

Qy   541 TGTAACAAGAATCGCAATTGAGGAGGCGGTCACAAAACAGTAAAGTGTTCCTGTCTACCTGGA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TGTAACAAGAATCGCAATTGAGGAGGCGGTCACAAAACAGTAAAGTGTTCCTGTCTACCTGGA 600

Qy   601 AAAGTGGGCTGGAACAACAAGAAAACCGGCCCTTCTTGGCTCGGATGCCTTCCATAGTGATTGGG 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 AAAGTGGGCTGGAACAACAAGAAAACCGGCCCTTCTTGGCTCGGATGCCTTCCATAGTGATTGGG 660

Qy   661 AAATGGGTGGTGTGAGATGGAGCCTTGCCTAGAGAGGAGGAGGAAGGAGGAAGAATGTAAGACACTCCCTGAC 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 AAATGGGTGGTGTGAGATGGAGCCTTGCCTAGAGAGGAGGAGGAAGGAGGAAGAATGTAAGACACTCCCTGAC 720

Qy   721 AATTCTGGATGGATGTCGGCAACAGGGCAACAAAATTAAGACCACGAGAAATTCACCAAGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AATTCTGGATGGATGTCGGCAACAGGGCAACAAAATTAAGACCACGAGAAATTCACCAAGA 780

Qy   781 ACCTAACAGAAGCGATTTGTTGGTAGTAAAGGAA 812
              ||||||||||||||||||||||||||||||||||
Db   781 ACCTAACAGAAGCGATTTGTTGGTAGTAAAGGAA 812

RESULT 2

ABQ77548
ID   ABQ77548 standard; cDNA; 2058 BP.
XX
AC   ABQ77548;
XX
DT   01-OCT-2002 (first entry)
XX
DE   Human phosphorylcholine/ethanolamine transferase 9.35-encoding cDNA.
XX
KW   Human; phosphorylcholine/ethanolamine transferase 9.35;
KW   recombinant production; gene therapy; lung hypectasia;
KW   bile acid metabolism disorder; angiocardiopathy; tumour; cancer;
KW   immune disorder; inflammatory condition; cytostatic; antiinflammatory;
KW   immunomodulator; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FH
FT   CDS   399..656
FT   /*tag= a
FT   /product= "Human phosphorylcholine/ethanolamine
FT   transferase 9.35"
XX
PN   CN1341736-A.
XX
PD   27-MAR-2002.
XX
PF   05-SEP-2000; 2000CN-00125017.
XX
PR   05-SEP-2000; 2000CN-00125017.
XX
PA   (SHAN-) SHANGHAI BIODOOR GENE DEV CO LTD.
XX
PI   Mao Y, Xie Y;
XX
DR   WPI; 2002-501219/54.
DR   P-PSDB; AAM49386.
XX
PT   New polypeptide-human phosphorylcholine/ethanolamine transferase 9.35 for
PT   treating lung hypectasia, bile acid metabolic disturbance disease,
PT   angiocardiopathy, related tumor, inflammation and immunological disease.
XX
PS   Claim 6; Page 26-27 (Disclosure); 34pp; Chinese.
XX
CC   The invention relates to human phosphorylcholine/ethanolamine transferase
CC   9.35 (AAM49386) and nucleic acids encoding it (ABQ77548). The protein has
CC   a molecular weight of 9.35 kD. The invention also relates to a method for
CC   the recombinant production of the protein, an antagonist of the protein,
CC   and the use of the protein, gene and antagonist in therapeutic
CC   applications. Phosphorylcholine/ethanolamine transferase 9.35 can be used
CC   in the treatment of a variety of diseases such as lung hypectasia,
CC   disorders of bile acid metabolism, angiocardiopathy, related tumors,
CC   immune disorders and inflammatory conditions. The present sequence
CC   represents cDNA encoding human phosphorylcholine/ethanolamine transferase
CC   9.35.
XX
SQ   Sequence 2058 BP; 609 A; 402 C; 437 G; 610 T; 0 U; 0 Other;

Query Match   84.5%; Score 686; DB 6; Length 2058;

RESULT 3
ABX71301
ID   ABX71301 standard; cDNA; 1936 BP.
XX
AC   ABX71301;
XX
DT   14-APR-2003 (first entry)
XX
DE   Human kidney-derived cDNA from clone DKFZphfkd2_4b6.
XX
KW   Human; gene; gene therapy; vaccine; disease treatment; detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO20112659-A2.
XX
PD   22-FEB-2001.
XX
PF   18-AUG-2000; 2000WO-IB001496.
XX
PR   18-AUG-1999; 99US-0149499P.
PR   28-SEP-1999; 99US-0155503P.
XX
PA   (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX
PI   Wiemann S;
XX
DR   WPI; 2001-327840/34.
DR   P-PSDB; ABU52838.
XX
PT   Nucleic acids having the sequences of clones isolated from libraries of
PT   different human tissues, useful in recombinant DNA methodologies.
XX
PS   Claim 1; Page 425; 1095pp; English.
XX
CC   This invention describes novel polynucleotides and polypeptides isolated
CC   from human cDNA libraries which can be used for gene therapy or in
CC   vaccines. The polynucleotides of the invention and antibodies encoded by
CC   them may be used in the prevention, diagnosis and treatment of diseases
CC   associated with inappropriate polypeptide expression. The products of the
CC   invention may also be used to identify modulators of expression and
CC   activity and to down regulate expression and activity. The antibodies of
CC   the invention may also be used as diagnostic agents for detecting the
CC   presence of polypeptides in samples. This sequence encodes a polypeptide
CC   described in the disclosure of the invention
XX
SQ   Sequence 1936 BP; 581 A; 371 C; 408 G; 576 T; 0 U; 0 Other;

Query Match   444; Conservative   96.1%;
Best Local Similarity 53.3%;   Score 433.2; DB 5; Length 1936;
Matches 444; Conservative 0; Pred. No. 1e-117; Mismatches 18; Indels 0; Gaps 0;

Best Local Similarity 89.4%;   Pred. No. 1e-192;
Matches 812; Conservative 0; Mismatches 0; Indels 96; Gaps 3;

Db   726 ACTCCCTGACAATTCTGGATGATGATGTGCGCAACAGGCAACAAAATTAAGACCACGAGAAT 785

Qy   771 TCACCCAAGAACCTAACAGAAGCATTTGTGGTAGTGAGTAAAGGAA 812
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   786 TCACCCAAGAACCTAACAGAAGCATTTGTGGTAGTGAGTAAAGGAA 827

RESULT 4
AAK51800
ID   AAK51800 standard; cDNA; 827 BP.
XX
AC   AAK51800;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human polynucleotide SEQ ID NO 345.
XX
OS   Homo sapiens.
XX
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
XX
PF   05-FEB-2001; 2001WO-US004098.
XX
PR   30-NOV-2000; 2000US-00728422.
PR   20-OCT-2000; 2000US-00693325.
PR   20-SEP-2000; 2000US-00663561.
PR   15-SEP-2000; 2000US-00654936.
PR   01-SEP-2000; 2000US-00620325.
PR   19-JUL-2000; 2000US-00598075.
PR   20-JUN-2000; 2000US-00560875.
PR   27-APR-2000; 2000US-00496914.
PR   03-FEB-2000;
XX
KW   nervous system disorder; arthritis; inflammation; ss.
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX
DR   P-PSDB; AAM78667.
DR   WPI; 2001-476283/51.
XX
PI   Xue AJ, Yang Y, Wejhrman T, Goodrich R;
PI   Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PA   Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
XX
XX
PT   (HYSE-) HYSEQ INC.
XX
PT   in diagnosis and gene therapy.
PT   Nucleic acids encoding polypeptides with cytokine-like activities, useful
XX
PS   Claim 1; Page 1351-1352; 622lpp; English.
XX
XX   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activity/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC   (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC   sequence listing were missing at the time of publication
XX
SQ   Sequence 827 BP; 230 A; 194 C; 198 G; 199 T; 0 U; 6 Other;
XX

Query Match   53.0%; Score 430.2; DB 4; Length 827;
Best Local Similarity   99.3%; Pred. No. 5.4e-117;
Matches 432; Conservative   0; Mismatches   3; Indels   0; Gaps   0;

Qy   378 TCAGGAAATGGCAAATGGTCTCTTGGCGATGTCCGATGTCCGGGGTCCTCGTGTATTTGTATTGTGGATTAAGTGCTTTG 437
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   210 TTATAGAATGGCAAATGGTCTCTTGGCGATGTCTTGGGGTCCTCGTGTATTTGTATTGTGGATTAAGTGCTTG 269

Qy   438 TGCAATGCTTACTCTGCCAATGGATCCCTTCAGCACACACACTTTCAGCACAGCATCACCTGCACAG 497
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   270 TGCAATGCTTACTCTGCCAATGGATCCCTTCAGCACACACACTTTCAGCACAGCATCACCTGCACAG 329

Qy   498 ACCAGAACGAGAGGGACGTGTGAAGTGATGTAGCACACGAGCATGTTGTAACAGAATCGCAT 557
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   330 ACCAGAACGAGAGGGACGTGTGAAGTGATGTAGCACACGAGCATGTTGTAACAGAATCGCAT 389

Qy   558 TGAGGAGCGGTCACAAACAGTAAAGTGTTCCGTGTCTACCTGGAAAAAGTGGCTGGAACAAC 617
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   390 TGAGGAGCGGTCACAAACAGTAAAGTGTTCCGTGTCTACCTGGAAAAAGTGGCTGGAACAAC 449

Qy   618 AAGAAACCGGCCTTCTTGGGTGGATGCCTCCATAGTGATTGGGAAAATGGTGGTGTGAGAT 677
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   450 AAGAAACCGGCCTTCTTGCGTCGGATGCCTCCATAGTGATTGGGAAAATGGTGGTGAGAT 509

Qy   678 GGAGCCTTGCCTAGGAGGAGAAGTGTAAGACACTCCCTGACAATTCTGGATGATGGATGTG 737
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   510 GGAGCCTTGCCTAGGAGGAGAAGAATGTAAGACACTCCCTGACAATTCTGGATGATGTG 569

Qy   738 CGCAACAGGCAACAAAAATTAAGACCACGAGAATTCACCCAAGAAGAACCTAACAGAAGCATTT 797
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   570 CGCAACAGGCAACAAAAATTAAGACCACGAGAATTCACCCAAGAAGAACCTAACAGAAGCATTT 629

Qy   798 GTGGTAGTGAGTAAAGGAA 812
         |||||||||||||||||||
Db   630 GTGGTAGTGAGTAAAGGGA 644

RESULT 5
AAK52784
ID   AAK52784 standard; cDNA; 671 BP.
XX
AC   AAK52784;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human polynucleotide SEQ ID NO 2313.
XX
OS   Homo sapiens.
XX
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
XX
PF   05-FEB-2001; 2001WO-US004098.
XX
PR   30-NOV-2000; 2000US-00728422.
PR   20-OCT-2000; 2000US-00693325.
PR   20-SEP-2000; 2000US-00663561.
PR   15-SEP-2000; 2000US-00654936.
PR   01-SEP-2000; 2000US-00620325.
PR   19-JUL-2000; 2000US-00598075.
PR   20-JUN-2000; 2000US-00560875.
PR   27-APR-2000; 2000US-00496914.
PR   03-FEB-2000;
XX
KW   nervous system disorder; arthritis; inflammation; ss.
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX
DE   Human polynucleotide SEQ ID NO 2313.
XX
XX
PT   (HYSE-) HYSEQ INC.
XX
PI   Xue AJ, Yang Y, Wejhrman T, Goodrich R;
PI   Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PA   Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
XX
DR   P-PSDB; AAM79651.
DR   WPI; 2001-476283/51.

XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
PS Claim 1; Page 4638; 6221pp; English.
XX
CC
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 671 BP; 182 A; 162 C; 167 G; 160 T; 0 U; 0 Other;

Query Match   52.4%; Score 425.4; DB 4; Length 671;
Best Local Similarity   98.6%; Pred. No.1.3e-115;
Matches   429; Conservative   0; Mismatches   6; Indels   0; Gaps   0;

Qy   378 TCAGAGAATGGCAATGGTCTCCTGGCGATGGTCCTGGGTCCTGTATTTGTGGATAAGTGCTTG 437
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   210 TTATAGAATGGCAAATGGTCTCCTGCGGATGTCTCCTGTATTGTGGATAAGTGCTTG 269

Qy   438 TGCAATGCTACTCTGCCATGGAATCCCTTCAGCACACTTTCAGCAGCATCACCTGCACAG 497
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   270 TGCAATGCTACTCTGCCATGGAATCCCTTCAGCACACTTTCAGCAGCATCACCTGCACAG 329

Qy   498 ACCAGAAGGAGGAGGAGACGTGTGAAGTGATTAGCACGACCACACATCCGAGTGTTGTTAACAAGAATGGCAT 557
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   330 ACCAGAAGGAGGAGGACGTGTGAAGTGATTAGCACGACCACACATCCGAGTGTTGTTAACAAGAATGGCAT 389

XX

PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PN 21-FEB-2000; 2000EP-00200610.
PF 26-FEB-1999;   99US-0122487P.
PR
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J; Duclert A; Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS Claim 1; SEQ ID NO 12748; 71pp + Sequence Listing; English.
XX
CC
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors
XX
SQ Sequence 261 BP; 70 A; 72 C; 52 G; 65 T; 0 U; 2 Other;

Query Match   23.0%; Score 187; DB 3; Length 261;
Best Local Similarity   98.9%; Pred. No. 4.7e-45;
Matches   187; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 CACTGGAGTGGGGATGGTCCATCGGCCAAACTATATAAACTGATTCTCATCAGAGAAACTGCACA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73 CACTGGANNGGGATGGTCCATCGGCAACATATAAACTGATTCTCATCAGAGAAACTGCACA 132

Qy   61 TTATCTCCCCATCCACTTCAAAAGGTCTTCGTCAGGCAGAGGTGACGCCAGGAGGATGATTAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   133 TTATCTCCCCATCCACTTCAAAAGGTCTTCGTCAGGCAGAGGTGACGCCAGGAGGATGATTAA 192

Qy   121 AGGTGAAAAATGACAAGGTTTCCACCCCTCAAAACCTTGGCTCTTCCTTTTCTGACAATACAGTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   193 AGGTGAAAAATGACAAGGTTTCCACCCCTCAAAACCTTGGCTCTTCCTTTTCTGACAATACAGTC 252

Qy   181 TGAATGAAC 189
              |||||||||
Db   253 TGAATGAAC 261

RESULT 6
AAC08673
ID   AAC08673 standard; cDNA; 261 BP.
XX
AC   AAC08673;
XX
DT   06-OCT-2000   (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 12748.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.

Qy   678 GGAGCCCTTGCCTAGGAGGAGAGAAGTAAGACACTCCCTGACAAATTCTGGATGGATGTG 737
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   510 GGAGCCCTTGCCTAGGAGGAGAGAAGTAAGACACTCCCTGACAAATTCTGGATGGATGTG 569

Qy   738 CGCAACAGGCACACAAAAATTAAGACCACCAGAATTCACCCAAGAACCTAACAGAGAAGCATTT 797
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   570 CGCAACAGGCACACAAAAATTAAGACCACCAGAATTCACCCAAGAACCTAACAGAGAAGGCTTT 629

Qy   798 GTGGTAGTAAAGGAA 812
              |||||||||||||||
Db   630 GTTGTAGTAAAGGGA 644

RESULT 7
AAI58347
ID   AAI58347 standard; cDNA; 716 BP.
XX
AC   AAI58347;
XX
DT   22-OCT-2001   (first entry)
XX
DE   Human polynucleotide SEQ ID NO 550.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

```
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US034263.
XX
PR   23-DEC-1999;            99US-00471275.
PI   21-JAN-2000;          2000US-00488725.
PI   25-APR-2000;          2000US-00552317.
PI   20-JUN-2000;          2000US-00598042.
PR   19-JUL-2000;          2000US-00620312.
PR   03-AUG-2000;          2000US-00633450.
PR   03-AUG-2000;          2000US-00653450.
PR   14-SEP-2000;          2000US-00662191.
PR   19-OCT-2000;          2000US-00693036.
PR   29-NOV-2000;          2000US-00723344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J,  Zhao QA;
PI   Zhou P,  Goodrich R,  Drmanac RT;
XX
PS   Claim 1; SEQ ID NO 550; 10078pp; English.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders such
PT   as central nervous system injuries.
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM39191.
XX
CC   The invention relates to human nucleic acids (AA157798-AA161369) and the
CC   encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders. Note: The sequence data for this patent did not form
CC   part of the printed specification.
XX
SQ   Sequence 716 BP; 212 A; 156 C; 195 G; 153 T; 0 U; 0 Other;

Qy            479 AGCAGCATCACCTGCACAGACCAGAAGCAGAGGACGAGCGTGTGAAGTGATAGCAGCACCGAT 538
                  ||||||||||||||||||||||||||||||||| || |||| || |||||  || |||||||
Db            295 ACCATCATAAAGCTCACCATTGTTAAAACGGGAGAACTTGTCTGTCTGTCTGCACTCCTACCTG 354

Qy            539 GTTGTAATAAGGAATCCGCATTGAGGAGCGGGTCACAAACAAGTAAAGTGTTCCTGTCTACCTG 598
                  | || ||||| |||  |||||  || || ||| ||||| || |  || |||||||||| |||||
Db            355 GCTGTAATAAGGAACAAGATAGAAGAACGGGTCACAAACAGTTGTGAGGTGGGCACTCACAGAT 414

Qy            599 GAAAAGTGGCTGGAACAAGAAAACCGGCCCTTCTTTGCGTCGGATGCCTCCATAGTGATTG 658
                  |||||||||||| ||||| |||||||||||| | | ||||||||| ||||||||||| | |
Db            415 GGCAGGTGGGCAGGACCACGCGAGCTGCTCCATCATGTGATGGATGCTTCACAACAGTGGAAC 474

Qy            659 CGAAATGGTGGGTGGTGTGAGATGGGAGGAGGAGAAGAATGTAAGACACTCCCTG 718
                  |||||||||||| ||||| |||||||||||||||||||||||||| |||||||||
Db            475 AGAAATGGTGGGTGCCTATATGCAGGCCATGTCTAGAGGGGAGGAAGAATGTAAGAGTTCTTCCGG 534
```

```
RESULT 8
ID   ADB48315 standard; cDNA; 716 BP.
AC   ADB48315;
XX
PN   US2003104529-A1.
XX
PD   05-JUN-2003.
XX
OS   Homo sapiens.
XX
KW   ss; cancer; neurodegenerative disease; human.
XX
DB   Novel human cDNA SEQ ID NO 225.
XX
DT   04-DEC-2003  (first entry)
AC   ADB48315;
XX
PF   04-JAN-2002; 2002US-00037270.
XX
PR   21-JAN-2000;          2000US-00488725.
PR   25-APR-2000;          2000US-00552317.
PR   19-JUL-2000;          2000US-00620312.
XX
PI   Zhou P,  Tang YT,  Liu C,  Asundi V,  Drmanac RT;
XX
DR   WPI; 2003-678194/64.
XX
PA   (ZHOU/)  ZHOU P.
PA   (TANG/)  TANG Y T.
PA   (LIU/)   LIU C.
PA   (ASUN/)  ASUNDI V.
PA   (DRMA/)  DRMANAC R T.
XX
PS   Claim 1; SEQ ID NO 225; 99pp; English.
XX
PT   New polynucleotide, useful for treating diseases e.g., cancer or
PT   neurodegenerative diseases.
XX
CC   The invention relates to a polynucleotide comprising a sequence given in
CC   the specification, or its mature protein-coding portion, or its
CC   complement. The polynucleotide is useful for treating diseases e.g.,
CC   cancer or neurodegenerative diseases and many others listed in the
CC   specification. The present sequence represents a novel human cDNA. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocId=20030104529.
XX
SQ   Sequence 716 BP; 212 A; 156 C; 195 G; 153 T; 0 U; 0 Other;

Qy            479 AGCAGCATCACCTGCACAGACCAGAAGCAGAGGACGAGCGTGTGAAGTGATAGCAGCACCGAT 538
                  ||||||||||||||||||||||||||||||||| || |||| || |||||  || |||||||
Db            295 ACCATCATAAAGCTCACCATTGTTAAAACGGGAGAACTTGTCTGTCTGTCTGCACTCCTACCTG 354

Qy            539 GTTGTAATAAGGAATCCGCATTGAGGAGCGGGTCACAAACAAGTAAAGTGTTCCTGTCTACCTG 598
                  | || ||||| |||  |||||  || || ||| ||||| || |  || |||||||||| |||||
Db            355 GCTGTAATAAGGAACAAGATAGAAGAACGGGTCACAAACAGTTGTGAGGTGGGCACTCACAGAT 414

Qy            599 GAAAAGTGGCTGGAACAAGAAAACCGGCCCTTCTTTGCGTCGGATGCCTCCATAGTGATTG 658
                  |||||||||||| ||||| |||||||||||| | | ||||||||| ||||||||||| | |
Db            415 GGCAGGTGGGCAGGACCACGCGAGCTGCTCCATCATGTGATGGATGCTTCACAACAGTGGAAC 474
```

Query Match   16.2%;   Score 131.2;   DB 4;   Length 716;
Best Local Similarity   63.1%;   Pred. No. 2.8e-28;
Matches 202;   Conservative   0;   Mismatches 118;   Indels   0;   Gaps   0;

Query Match   16.2%;   Score 131.2;   DB 8;   Length 716;
Best Local Similarity   63.1%;   Pred. No. 2.8e-28;
Matches 202;   Conservative   0;   Mismatches 118;   Indels   0;   Gaps   0;

```
Qy            719 ACAATTCTGGATGATGTGCGCAACAGGCAACAAAATTAAGACCACGAGAATTCACCAA 778
                  |||| | ||||||| | |||| ||||| || ||||||||||  || || || |||| ||
Db            535 ATCGGAAAGGATGGAGCTGTTCCTCTGGGATAAAAGTCAAAACAACTAGGGTAACCATT 594

Qy            779 GAACCTAACACAAGCATTTG 798
                  |||||| || |||||||||
Db            595 AACCCAGGAGAATCAAGTG 614
```

Db   GGCAGGTGGGCCAGGCGACGGCACGGCGAGCTGCTCCATCATGTGTGGATGCTTCAATAGTGGAAC
Qy   415 ... 474
Qy   659 GGAAATGTGGTGTGAGATGGAGCCTTGCCTAGGAGAGAAGAGAAGATGAAGACACTCCCTG 718
Db   475 AGAAAATGGTGGTGCCATARGACAGCCCATGTCTACAGGGAGAGAATGTAAAGTTCTTCCG 534
Qy   719 ACAATTCTGGATGGATGTGCGCAACAGGCCAACAAAATTAAGACCACGAGAATTCACCCAA 778
Db   535 ATCGGAAAGGATGGAGCTGTTCCTCTGGGAATAAAGTCAAAACACTAGGGTAACCCATT 594
Qy   779 GAACCTAACAGAACGCATTTG 798
Db   595 AACCCAGGAGAGAATCAAGTG 614

RESULT 9
AAI60133
ID   AAI60133 standard; cDNA; 1171 BP.
XX
AC   AAI60133;
XX
DT   22-OCT-2001 (first entry)
XX
DE   Human polynucleotide SEQ ID NO 4122.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PD   26-DEC-2000; 2000WO-US034263.
XX
PR   23-DEC-1999;   99US-00471275.
PR   21-JAN-2000; 2000US-00488725.
PR   25-APR-2000; 2000US-00552317.
PR   20-JUN-2000; 2000US-00598042.
PR   19-JUL-2000; 2000US-00620312.
PR   03-AUG-2000; 2000US-00653450.
PR   14-SEP-2000; 2000US-00662191.
PR   19-OCT-2000; 2000US-00693036.
PR   29-NOV-2000; 2000US-00727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI   Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI   Zhou P, Goodrich R, Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM40977.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders such
PT   as central nervous system injuries.
XX
PS   Claim 1; SEQ ID NO 4122; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC   encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies, and central nervous system diseases, such as
CC   localised neuropathies, and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders. Note: The sequence data for this patent did not form
CC   part of the printed specification
XX
SQ   Sequence 1171 BP; 329 A; 265 C; 321 G; 254 T; 0 U; 2 Other;

Query Match   16.2%; Score 131.2; DB 4; Length 1171;
Best Local Similarity   63.1%;
Matches 202; Conservative   0; Mismatches 118; Indels   0; Gaps   0;

Qy   479 AGCAGCATCACCTTCACAGACCAGAAGAGAAGAAGGAGGAGGACGTGTGAAGTGATAGAGCACGGAT 538
Db   673 ACCATCAATAAAGTCTCACCCATGTTAAACGGGGAACTTGTTGAGGTGGGTGGCACTCCAGAGAT 732
Qy   539 GTTGTAACAAGAATCGCATTGTGAGGAGGCTGCTCAAAACAGTAAAGTGTTCCTTGTCTTACCTG 598
Db   733 GCTGTAATAAGGACAAGAATAGAAGGAAGAACGGGTCACGAAACAGTCAAGTGCTGTCCGCGTTCCCTG 792
Qy   599 GAAAAGTGGCTGGACAACAACAAGAAAACCGGGCCTTCTTTGCGGTCGATGCCTCCAATAGTGATTG 658
Db   793 GGCAGGTGGGCCAGGCGACGGCACGGCGAGCTGCTCCATCATGTGTGGATGCTTCAATAGTGGAAC 852
Qy   659 GGAAATGTGGTGTGAGATGGAGCCTTGCCTAGGAGAGAAGAGAAGATGAAGACACTCCCTG 718
Db   853 AGAAAATGGTGGTGCCATAGGACAGCCCATGTCTACAGGGAGAGAATGTAAAGTTCTTCCG 912
Qy   719 ACAATTCTGGATGGATGTGCGCAACAGGCCAACAAAATTAAGACCACGAGAATTCACCCAA 778
Db   913 ATCGGAAAGGATGGAGCTGTTCCTCTGGGAATAAAGTCAAAACACTAGGGTAACCCATT 972
Qy   779 GAACCTAACAGAACGCATTTG 798
Db   973 AACCCAGGAGAGAATCAAGTG 992

RESULT 10
ADB62738
ID   ADB62738 standard; cDNA; 2390 BP.
XX
AC   ADB62738;
XX
DT   04-DEC-2003 (first entry)
XX
DE   Human cDNA encoding clone NT2NE20069580.
XX
KW   Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW   tissue regeneration; cell regeneration; membrane protein;
KW   signal transduction-related protein; transcription-related protein;
KW   osteoporosis; neurological disease; cancer; tumour.
XX
OS   Homo sapiens.
XX
PN   EP1308459-A2.
XX
PD   07-MAY-2003.
XX
PF   28-MAR-2002; 2002EP-00007401.
XX
PR   05-NOV-2001; 2001JP-00379298.
PR   25-JAN-2002; 2002US-00350978.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
XX
FH   Key   Location/Qualifiers
FT   CDS   1404..2282
FT   /tag= a
FT   /product= "Clone NT2NE20069580 protein"

PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-450961/43.
DR P-PSDB; ADB64708.
XX
PT New polynucleotides and polypeptides, useful for developing a diagnostic
PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.
XX
PS Claim 1; Page; 222pp; English.
XX
CC The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesising the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours). The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a cDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.
XX
SQ Sequence 2390 BP; 524 A; 629 C; 527 G; 710 T; 0 U; 0 Other;

Query Match   15.8%; Score 128.4; DB 9; Length 2390;
Best Local Similarity   92.5%; Pred. No. 3.3e-27;
Matches 135; Conservative   0; Mismatches 11; Indels   0; Gaps   0;

Qy 357 CTCTGCTTCTTGGCTGGAATTTTCAGGAGAATGGCAATGGTCTCTGGAATGGTCCTGGGTCCT 416
Db 610 CTTTCCTCGTCGGAAATGTTTCTTCTTTAGAGAATGGCAATGGTCTTCTCCGGATGGTCCTGGGGTCCT 669

Qy 417 GTATTTGTGGATAAGTGCTTGTGCAATGCTACTCTTCCCATGGATCCCTTCAGCACACTTT 476
Db 670 GTATTTGTGGATAAGTGCTTGTGCAATGCTACTCTTCCCATGGATCCCTTCAGCACACTTT 729

Qy 477 CCAGCAGCATCACCTGGCACAGACCAG 502
Db 730 CCAGCAGCAATCACCTGCACAGACCAG 755

RESULT 11
AAC25659
ID AAC25659 standard; cDNA; 238 BP.
AC AAC25659;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 29734.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.

XX
PF 21-FEB-2000; 2000EP-00200610.
XX
PR 26-FEB-1999; 99US-0122487P.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
    Claim 1; SEQ ID NO 29734; 71pp + Sequence Listing; English.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 238 BP; 69 A; 47 C; 72 G; 50 T; 0 U; 0 Other;

Query Match   13.5%; Score 109.6; DB 3; Length 238;
Best Local Similarity   69.8%; Pred. No. 4.2e-22;
Matches 162; Conservative   0; Mismatches 69; Indels   1; Gaps   1;

Qy 503 AAGGAGGGGACGTGTGAAGTGATAGCAGCACCACCGGATTTGTAACAAGAATCGGCATTGAGG 562
Db   2 AAACGGGGAACTTGTGAGGTGGTGGCACTCCACAGATGCGTTGTAATAAGAACAAGATAGAAG 61

Qy 563 AGCGGTCACAAACAGTAAAGTGTTCCTGTCTACCTGGAAAAGTGGCTGGAACAACAAGAA 622
Db   62 AACGGTCACAAACAGTCAAGTGCTTCCTGCTTCCCTGGGCAGGTGGCACACGCGGAG 121

Qy 623 ACCG-GCCTTCTTGCGTCGGATGCCTCCATAGTGAATTGGGAAATGGGTCGGTGTTGAGGATGGAG 681
Db 122 GCTGCGTCCATCATGTGGCATGGATGCTTCAATAGTGAACACAGAAAACAGGGTGGTGCCATATGCAG 181

Qy 682 CCTTGCCTAGAAGGAGGAGAAGAAATGTTAAGACACTCCCTGACAAATTCTGGATTGGA 733
Db 182 CCAATGCTAGAGAGGGGAGAAGAATGTTAAAGTTCTTTCCGGATCGGAAAGGAATGGA 233

RESULT 12
AAX51967
ID AAX51967 standard; DNA; 333 BP.
AC AAX51967;
XX
DT 22-JUN-1999 (first entry)
XX
DE Human secreted protein 5' EST SEQ ID NO: 181.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS Homo sapiens.

XX
PN   WO9906552-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998; 98WO-IB001236.
XX
PR   01-AUG-1997; 97US-0005223.
XX
PA   (GENSET) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Lacroix B;
XX
DR   WPI; 1999-153782/13.
DR   P-PSDB; AAY13167.
XX
PT   New isolated brain-derived nucleic acids - used to develop products which
PT   may have cytokine, immune, regulatory, haematopoiesis regulating, anti-
PT   inflammatory or tumour inhibition activity.
XX
PS   Claim 1; Page 353; 577pp; English.
XX
CC   AAX51787 to AAX52019 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY12987 to
CC   AAY13219, respectively. The proteins given represent the signal peptide
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell
XX
SQ   Sequence 333 BP; 66 A; 95 C; 80 G; 92 T; 0 U; 0 Other;

Query Match   10.9%; Score 88.8; DB 2; Length 333;
Best Local Similarity   97.8%; Pred. No. 7.4e-16;
Matches   90; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   378 TCAGAGAAATGGCAAATGGTCTCTGCGGATGTCTGGGGTTCCTGTTATTTGTGGATAAGTGCTTG 437
             ||||||||||||||||||||| ||||||| |||||||| |||||||||||||||||||||
Db   240 TTATAGAAATGGCAAATGGTCTCTCCGATGTCCTGGGTCCTGTATTTGTGGATAAGTGCTTG 299

Qy   438 TGCAATGCTACTCTGCCATGGATCCCTTCAGC 469
             ||||||||||||||||||||||||||||||||
Db   300 TGCAATGCTACTCTGCCATGGATCCCTTCAGC 331

RESULT 13
ID   AAH99019
AAH99019 standard; cDNA; 429 BP.
AC   AAH99019;
XX
DT   12-OCT-2001 (first entry)
XX
DE   Human EST-derived coding sequence SEQ ID NO: 876.
XX
KW   Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW   tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW   diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;
KW   gene therapy; nutrition; ss.
XX
OS   Homo sapiens.
XX
PN   WO200154477-A2.

XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US002687.
XX
PR   25-JAN-2000; 2000US-00491404.
PR   17-JUL-2000; 2000US-00617746.
PR   03-AUG-2000; 2000US-00631451.
PR   15-SEP-2000; 2000US-00663870.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI   Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR   WPI; 2001-476164/51.
DR   P-PSDB; AAM24360.
XX
PT   Isolated polypeptide for treatment of diseases, diagnostics, raising
PT   antibodies and research use.
XX
PS   Claim 1; Page 711; 1275pp; English.
XX
CC   The present invention provides the protein and coding sequences of novel
CC   proteins from a variety of organisms, including human, dog, cat, horse,
CC   cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC   urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC   from the organism of interest. They can be used in diagnostics,
CC   forensics, gene mapping, identification of mutations, to assess
CC   biodiversity and for nutritional purposes. The present sequence is a cDNA
CC   of the invention
XX
SQ   Sequence 429 BP; 80 A; 132 C; 131 G; 86 T; 0 U; 0 Other;

Query Match   8.6%; Score 70; DB 4; Length 429;
Best Local Similarity   65.2%; Pred. No. 3.2e-10;
Matches 103; Conservative   0; Mismatches 55; Indels   0; Gaps   0;

Qy   508 GGGACGTGTGAAGTGATGATAGCAGCACACCGGATGTTGTTAACAAGAATCGCATTGAGGAGCGG 567
             || ||| |||||||||| |||| | ||||| ||||| || | |||||| ||| |||||||||
Db   213 GGCACCTGCGAGGTGATTGCGGCTCCACCGGTGCTCGCTGCAACCGGAACCGCATCGAGGAGCGC 272

Qy   568 TCACAAACACAGTAAAAGTGTTCCTGTCTCTACCTGGAAAAAGTGGCTGGAACAACAAGAAAACCGG 627
             ||| | ||| || |||| ||| ||| | ||| ||||| |||| |||| |||| ||||
Db   273 TCCCAGACAGGTGAAAATGCTCCGTGTTTTTTCTGGCCAGGTGGCGCGCCACCACGCGGGGCAAAG 332

Qy   628 CCTTCTTGGCGTCGGATGCCCTCCAATAGTGATTGGGAAATG 665
             || |||| ||||| || || | ||| |||| |||||
Db   333 CCCTCCTGCGTGGGACGACCTGCTCTTGGCGTGCCCACTG 370

RESULT 14
ID   AAZ33334
AAZ33334 standard; cDNA; 1493 BP.
AC   AAZ33334;
XX
DT   29-FEB-2000 (first entry)
XX
DE   Human secreted protein clone ml243_1 nucleotide sequence SEQ ID NO:37.
XX
KW   Human; secreted protein; nutritional; cytokine; cell proliferation;
KW   differentiation; immune stimulating; vaccine; suppression;
KW   haematopoiesis regulation; tissue growth; activin; inhibin; chemotactic;
KW   chemokinetic; haemostatic; thrombolytic; receptor; ligand;
KW   anti-inflammatory; cadherin; tumour invasion suppressor;
KW   tumour inhibition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9957132-A1.
XX
PD   11-NOV-1999.

XX
PF   07-MAY-1999;
XX
PR   07-MAY-1998; 99WO-US009970.
PR   02-JUN-1998; 98US-0084564P.
PR   22-JUL-1998; 98US-0087645P.
PR   31-JUL-1998; 98US-0097212P.
PR   10-AUG-1998; 98US-0094935P.
PR   11-AUG-1998; 98US-0095880P.
PR   06-MAY-1999; 98US-0096068P.
                   99US-00306111.
XX
PA   (GEMY) GENETICS INST INC.
XX
PI   Jacobs K, Mccoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI   Merberg D, Treacy M, Agostino MJ, Steininger RJ, Bowman MR;
PI   Diblasio-Smith E, Widom A;
XX
DR   WPI; 2000-052937/04.
DR   P-PSDB; AAY53016.
XX
PT   New polynucleotides encoding secreted human proteins, derived from adult
PT   placenta, adult retina, fetal brain, fetal.
XX
PS   Claim 46; Page 396-397; 492pp; English.
XX
CC   The present invention describes new human secreted proteins which were
CC   isolated from adult placenta, adult retina, foetal brain, foetal kidney,
CC   adult blood, adult brain, adult thyroid, adult bladder, adult neural
CC   tissue, adult testes, and adult lymph node cDNA libraries. The human
CC   secreted proteins, and the polynucleotides encoding them, are predicted
CC   to have biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals. Suggested activities include nutritional activity, cytokine and
CC   cell proliferation/differentiation activity, immune stimulating (e.g. as
CC   vaccines) or suppressing activity, haematopoiesis regulating activity,
CC   tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC   receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC   invasion suppressor activity, and tumour inhibition activity. The
CC   polynucleotides are also stated to be useful for gene therapy. AAZ33316
CC   to AAZ33373 encode human secreted proteins, and AAY52998 to AAY53060
CC   represent human secreted proteins, given in the present invention
XX
SQ   Sequence 1493 BP; 299 A; 476 C; 455 G; 262 T; 0 U; 1 Other;

Query Match   51.5%; Score 52.2; DB 3; Length 1493;
Best Local Similarity   51.5%; Pred. No. 0.00011;
Matches 120; Conservative   0; Mismatches 113; Indels   0; Gaps   0;

Qy   508 GGGACGTGTGAAGTGATAGCAGCACACCGATGTTGTAACAAGAATCGCCATTGAGGACGG 567
Db   230 GGCACCGTGTGAGATTGTGACCTTGGACCCAGCCTTCGGAGGACGATCGCC 289
Qy   568 TCACAAACAGTAAAGTGTTCCTGTCTACCTGGAAAAAGTGGCTGGAACAACAAGAAACCGG 627
Db   290 CGGCAGACCGCCCGCTGTGCGTGTTAGAAAGGGCAGATCGCCCGCGCCACCGAGAGCCCGG 349
Qy   628 CCTTCTTGCGTCGATGCCTCCATAGTGATTGGGAAATGGTGGTTGTGATGGAGGAGCCTTGC 687
Db   350 CCCGCCCTGTGTGACGCAAGAAACATCAAGACCAGCTGGTGTGTGACATGCTTCCGTGT 409
Qy   688 CTAGAAGGAAGAAATGTAAGACACACTCCCTGACAAATTCTGGATTGGATGATGTGCGC 740
Db   410 CTGGAAGGGGGAAGGCTGCGCACTTGTTAATCAACCGGTCAGGCTGGAGCGTGCAC 462

XX
DE   Human secreted protein encoding cDNA for gene 42.
XX
KW   Secreted protein; human; immunosuppressive; antiproliferative; antiarthritic;
KW   antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW   nootropic; neuroprotective; antibacterial; virucide; fungicide;
KW   ophthalmological; gene therapy; treatment; autoimmune disease; infection;
KW   cardiovascular disorder; nervous system disorder; ocular disorder;
KW   wound healing; epithelial cell proliferation; skin aging; mental state;
KW   transplantation; metabolism modulation; ss.
XX
OS   Homo sapiens.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Ruben SM, Komatsoulis G;
XX
PN   WO200055200-A1.
XX
PD   21-SEP-2000.
XX
DR   WPI; 2000-656008/63.
DR   P-PSDB; AAB44958.
XX
PF   09-MAR-2000; 2000WO-US0060042.
XX
PR   12-MAR-1999; 99US-0124143P.
PR   03-DEC-1999; 99US-0168663P.
XX
PT   Isolated human secretory proteins, nucleic acids encoding them and
PT   antibodies directed against them, useful for diagnosing and treating
PT   disorders related to the proteins such as cancer, Alzheimer's disease and
PT   Parkinsons.
XX
PS   Claim 1a; Page 373-374; 453pp; English.
XX
CC   This invention describes a novel isolated polypeptide (I) and its
CC   encoding nucleic acid molecule (II) which have immunosuppressive,
CC   antiproliferative, antirheumatic, antiproliferative, cytostatic, cardiant,
CC   antiarthritic, antirheumatic, antiproliferative, cytostatic, cardiant,
CC   vasotropic, cerebroprotective, nootropic, neuroprotective, antibacterial,
CC   virucide, fungicide and ophthalmological activity and which can be used
CC   for gene therapy. (I) and (II) are used to prevent, treat or ameliorate a
CC   medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
CC   dogs, chickens or sheep. (I) and (II) are also used in diagnosing a
CC   pathological condition or susceptibility to a pathological condition. The
CC   antibodies to (II) can also be used in alleviating symptoms associated
CC   with the disorders and in diagnostic immunosorbent assays e.g. radioimmunoassays
CC   or enzyme linked immunosorbent assays (ELISA). Disorders which are
CC   diagnosed or treated include autoimmune diseases e.g. rheumatoid
CC   arthritis; hyperproliferative disorders e.g. neoplasms of the breast or
CC   liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
CC   disorders e.g. cerebral ischemia, angiogenesis, nervous system disorders
CC   e.g. Alzheimer's disease, infections caused by bacteria, viruses and
CC   fungi and ocular disorders e.g. corneal infection. The polypeptides can
CC   also be used to aid wound healing and epithelial cell proliferation, to
CC   prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides are used to
CC   modulate mammalian metabolism, to change mammal's mental state or
CC   physical state by influencing biorhythms circadian rhythms, depression
CC   tendency for violence tolerance for pain, reproductive capabilities,
CC   hormonal or endocrine levels, appetite, libido, memory, stress or other
CC   cognitive qualities, as a food additive or preservative, such as to
CC   increase or decrease storage capabilities, fat content, lipid, protein,
CC   carbohydrates, vitamins, minerals, cofactors or other nutritional
CC   components
XX
SQ   Sequence 2469 BP; 522 A; 709 C; 681 G; 550 T; 0 U; 7 Other;

Query Match   51.5%; Score 52.2; DB 3; Length 2469;
Best Local Similarity   51.5%; Pred. No. 0.00014;
Matches 120; Conservative   0; Mismatches 113; Indels   0; Gaps   0;

RESULT 15
ID   AAC79939 standard; cDNA; 2469 BP.
XX
AC   AAC79939;
XX
DT   12-FEB-2001 (first entry)

```
Qy   508 GGGACCGTGTGAAGTGATAGCAGCACACCGATGTTGTAACAAGAATCGCATTGAGGAGCGG 567
         || |||||  |||||  | |       |  |     |  |          |   |
Db   121 GGCACCTGTGAGATTGTTGACCTTGGACCGGGGGACAGCGACGCCTCGGGAGGACGATCGCC 180

Qy   568 TCACAAACAGTAAAGTGTTCCTGTCTACCTGGAAAAGTGGCTGGAACACACAAGAAACCGG 627
         | |   | | |  |  |||||| || ||  |    |||     |  |       || ||||
Db   181 CGGCAGACCGCCCGCTGTGCCGTGTAGAAAGGGGCAGATCGCCGCACCACGAGAGCCCGG 240

Qy   628 CCTTCTTGCGTCGATGCCTCCAATAGTGAATTGGGAAATGGTGGGTGTGAGATGGAGCCTTGC 687
         || |||  |       ||  |||   || | |||||||||||||||| ||||||||||||||
Db   241 CCCGCCTGTGTGGGACCGCAAGAAATCATCAAGACACCAGTGGGTGTGACATGCTTCCGTGT 300

Qy   688 CTAGAAGGAAGAAATGTAAGACACTCCCTGACAATTCTGGATGGGATGTGCGC 740
         |  |||  | |   |      | | ||  |||| | ||  ||   | ||| |||
Db   301 CTGGAGGGGGGAGGAAGGCTGCGACTTGTTAATCAACGGGTCAGGCTGGACGGTGCAC 353
```

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: April 9, 2004, 09:39:26 ; Search time 3291 Seconds
                 (without alignments)
                 10694.179 Million cell updates/sec

Title: US-09-763-335-4
Perfect score: 812
Sequence: 1 cactggagtggggatggtcc..........catttgtggtagtaaaggaa 812

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151699S residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
                 1: gb_ba:*
                 2: gb_htg:*
                 3: gb_in:*
                 4: gb_om:*
                 5: gb_ov:*
                 6: gb_pat:*
                 7: gb_ph:*
                 8: gb_pl:*
                 9: gb_pr:*
                 10: gb_ro:*
                 11: gb_sts:*
                 12: gb_sy:*
                 13: gb_un:*
                 14: gb_vi:*
                 15: gb_ba:*
                 16: em_fun:*
                 17: em_hum:*
                 18: em_in:*
                 19: em_mu:*
                 20: em_om:*
                 21: em_or:*
                 22: em_ov:*
                 23: em_pat:*
                 24: em_ph:*
                 25: em_pl:*
                 26: em_ro:*
                 27: em_sts:*
                 28: em_un:*
                 29: em_vi:*
                 30: em_hum:*
                 31: em_htg_inv:*
                 32: em_htg_mus:*
                 33: em_htg_other:*
                 34: em_htg_pln:*
                 35: em_htg_rod:*
                 36: em_htg_mam:*
                 37: em_htg_vrt:*
                 38: em_sy:*
                 39: em_htgo_hum:*
                 40: em_htgo_mus:*
                 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Strand	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	450.2	55.4	801	10	AY325120	AY325120 Mus muscu
	2	433.2	53.3	1936	6	AX086349	AX086349 Sequence
	3	433.2	53.3	1936	6	HSM803068	AL713702 Homo sapi
c	4	431.8	53.2	815	9	AY325114	AY325114 Homo sapi
c	5	368.8	45.4	137703	9	AC092036	AC092036 Homo sapi
c	6	368.8	45.4	158080	9	AC138390	AC138390 Homo sapi
	7	368.8	45.4	187748	9	AC139766	AC139766 Homo sapi
c	8	215.2	26.5	247515	2	AC116080	AC116080 Rattus no
c	9	190.2	23.4	193027	10	AC117234	AC117234 Mus muscu
c	10	187	23.0	261	6	AX896885	AX896885 Sequence
	11	187	23.0	261	6	BD032418	BD032418 Sequence
	12	145	17.9	172288	9	AC104167	AC104167 Homo sapi
	13	145	17.9	173990	9	AC112214	AC112214 Homo sapi
	14	137.4	16.9	901	10	AY325123	AY325123 Mus muscu
	15	136.8	16.8	1897	9	AK057890	AK057890 Homo sapi
	16	136.8	16.8	2188	9	BC031566	BC031566 Homo sapi
	17	136.8	16.8	2274	9	AY325117	AY325117 Homo sapi
	18	135.2	16.7	801	10	AY325122	AY325122 Mus muscu
	19	134.2	16.5	2011	9	AB063062	AB063062 Macaca fa
	20	134.2	16.5	2265	9	AB045997	AB045997 Macaca fa
	21	132.6	16.3	1077	9	AY325116	AY325116 Homo sapi
	22	131.2	16.2	671	6	AY325115	AY325115 Homo sapi
	23	131.2	16.2	716	6	AR338734	AR338734 Sequence
	24	131.2	16.2	2718	9	AK123580	AK123580 Homo sapi
c	25	131.2	16.2	3474	9	AC102744	AC102744 Mus muscu
	26	131.2	16.2	4058	9	BC050347	BC050347 Homo sapi
	27	128.4	15.8	2390	9	AX747367	AX747367 Sequence
	28	128.4	15.8	2390	9	AK092129	AK092129 Homo sapi
c	29	128	15.8	3087	9	BC028403	BC028403 Homo sapi
	30	127.6	15.7	43205	9	AC126123	AC126123 Homo sapi
	31	124	15.3	1101	10	AY325121	AY325121 Mus muscu
	32	124	15.3	4039	10	BC027082	BC027082 Mus muscu
c	33	122.6	15.1	243881	2	AC135662	AC135662 Rattus no
c	34	122.6	15.1	207096	2	AC102744	AC102744 Mus muscu
c	35	109.6	13.5	238	6	AX913871	AX913871 Sequence
c	36	109.6	13.5	238	6	BX890603	BX890603 Danio rer
c	37	100.8	12.4	158186	2	BD049404	BD049404 Sequence
c	38	100.8	10.9	333	10	BX890603	BX890603 Danio rer
c	39	88.2	10.9	179979	2	BD077927	BD077927 5'EST of
c	40	80.2	9.9	137393	2	BX571952	BX571952 Danio rer
c	41	80.2	9.9	163795	2	BX545852	BX545852 Danio rer
c	42	79.4	9.8	218896	2	AC106975	AC106975 Rattus no
c	43	77.2	9.5	171705	9	AC104166	AC104166 Homo sapi
c	44	77.2	9.5	218652	9	AC096922	AC096922 Homo sapi
c	45	76.8	9.5	212238	2	AC134077	AC134077 Rattus no

ALIGNMENTS

RESULT 1
AY325120
```
LOCUS       AY325120        801 bp    mRNA    linear   ROD 23-JUL-2003
DEFINITION  Mus musculus TAFA1 mRNA, complete cds.
ACCESSION   AY325120
VERSION     AY325120.1  GI:32967238
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 801)
AUTHORS     Tang,Y.T., Emtage,P., Funk,W., Hu,T., Arterburn,M., Park,E. and
            Rupp,F.
TITLE       TAFA: A Novel Secreted Family with Homology to CC-chemokines
```

JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 801)
  AUTHORS   Tang,Y.T., Emtage,P., Funk,W., Hu,T., Arterburn,M., Park,E. and
            Rupp,F.
  TITLE   Direct Submission
  JOURNAL   Submitted (17-JUN-2003) Biology Research, Nuvelo, 675 Almanor
            Avenue, Sunnyvale, CA 94085, USA
FEATURES   Location/Qualifiers
     source   1..801
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
     CDS   238..639
                     /note="secreted protein similar to CC-chemokines"
                     /codon_start=1
                     /product="TAFa1"
                     /protein_id="AAP92412.1"
                     /db_xref="GI:32967239"
                     /translation="MAMYSAMSWALYLWISACAMLLCHGSLQHTFQOHHLHRPEGGTC
            EVTAAHRCCNKRPIEERSQTVKCSCLPGKVAGTTRNRPSCVDASIVIGKWWCEMEPCL
            EGEBCKTLPDNSGMWCATGNKIKTRIHPRT"
ORIGIN
Query Match   55.4%; Score 450.2; DB 10; Length 801;
Best Local Similarity   86.4%; Pred. No. 8.5e-109;
Matches 497; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy   238 TTTTTTTTTTAATCTGATAAAGAAGAAGATTGTTGGGAAGCTCTTTGAAAAAAATTTTAAA 297
          |||||||||| |||| ||||||||| ||| ||||||||||||| |||||||||| ||||
Db   91 TTTTTTTTTCAATTCTTAATAAAGAATTGCTGGGAAGCATTCTTCTTTGAAAAATCTCAGAA 150

Qy   298 TTGTGGCACAGATGGATTTTAAATAAGTGTTAGATCTTTCCAAATGAACACTAATAGAGTAC 357
          ||||||||||||||||||||||| |||||||| |||||||||||||| ||||| |||| ||
Db   151 CTGTGGCACAGATGGATTTTAAAAAGTGTTAGCTCTTTCCAAATGAAGACACTAGGAGGGTTC 210

Qy   358 TCTGCTCTTGGCTGGATTTTTCAGAGAATGGCAATGGTCTCTGCCAATGTCCTGCCTG 417
          ||||||||||||||||||||||||||||||||||||||||| |||||||||| ||||
Db   211 CCTGCTCTTGGCTGGATTTTTCAGAGAATGGCAATGGTCTCTTGCAATGTCCTGGGCCCTG 270

Qy   418 TATTTGTTGGATAAAGTTGGCTTGGTCATCTCACTCCCTTCAGCACACTTTG 477
          || ||||||||||| || |||| || |||| ||||| | ||| ||| ||||
Db   271 TACTTGTTGGATAAGTGCTTGTCGATGGCTGTCTCTCGCCACTCCAACACCTTC 330

Qy   478 CAGCAGCATCACCTGCAACAGACAGAAGCAGAAGGAGGAGCGTGTGACATAGCACACCGA 537
          |||||||||||||||| ||| || |||| ||||||||||||||||| | |||| ||||
Db   331 CAGCAGCATCACCTGCGACCGGCAGAAGGAGGAGCGTGTGATCGGCAACACACAGG 390

Qy   538 TGTTGTAACAAGAAATGCCATTGAGGAGGAGCGTAAAGTGTTTTCGTTGTCTACCT 597
          |||||||||||||| || ||| ||||||||||||| ||||| || ||| ||||||
Db   391 TGTTGTAACAAGAAACGCATCGAGGAGGAGCGTCACAAACAGTAAGTGTTTCCTTGTTTACCT 450

Qy   598 GGAAAAGTGGCTCGGAACAACAAGAAAACCGACCGCTTCTTCGGCCTCCCATCCATAGTGATT 657
          || |||||||| || ||||||||||| |||||| ||| | ||| | ||||||||| |||||
Db   451 GGGAAAGTGGCTGGGGACAACAAGAACCGACCTTCCTGTGTGGGATCCCTCCATAGTAATT 510

Qy   658 GGGAAAATGGTGGTGTGAGAATGGAGCCTTGCTCTAGAGAGAGAGAGAATGTAAGACACTCCCT 717
          |||||| ||||||| ||||||||||||| ||| || || ||||| |||||||||||||||||||
Db   511 GGGAAATGGTGGGTGTGAGAATGGAGCCCTGCCTAGAAGGAGAAGAGAATGTAAGACACTCCCT 570

Qy   718 GACAAATTCTGGATTGGATGGCGCACACAGGCCAACACAAAATTAAGACCACGAGAATTCACCCA 777
          ||||| |||||||| |||||| | | | ||||| |||||| |||||||| || |||||||||||||
Db   571 GACAATTCTGGATGGATGGTGTGCCTACAGGCAACACAAGATTAAGACTACACAGAATTCACCCA 630

Qy   778 AGAACCTAACACAGAAGCATTGTTGGTAGTAAAGGAA 812
          ||||||||||||||||||||||||| || || ||||
Db   631 AGAACCTAACACAGAAGCATTGTTGTATATAAAGGAAAGGA 665

ACCESSION   AX086349
VERSION   AX086349.1 GI:13275945
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Wiemann,S.
  TITLE   Human dna sequences
  JOURNAL   Patent: WO 0112659-A 301 22-FEB-2001;
            German Human Genome Project (DE)
FEATURES   Location/Qualifiers
     source   1..1936
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN
Query Match   53.3%; Score 433.2; DB 6; Length 1936;
Best Local Similarity   96.1%; Pred. No. 2.9e-104;
Matches 444; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy   351 AGAGTACTCTGCTCTTGGCTGGATTTTTCAGAGAGAATGGCAATGGTCTCTGCCGATGTCCTG 410
          || ||| |||||| | | | ||| ||||||| |||||||||||||||||| ||||||||||
Db   366 AGGTAACCTTTCCTCGTGCTGAAATGTTTCTTTAGAGAATGGCAATGGTCTCTTGCCGATGTCCTG 425

Qy   411 GGTCCTGTTATTGTTGGATAAGTGCTTGTCGATGGCTACTCTGCCATGGATCCCTTCAGCA 470
          |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   426 GGTCCTGTTATTTGTGGATAAGTGCTTGTCGATGGCTACTCTGCCATGGATCCCTTCAGCA 485

Qy   471 CACTTTCCAGCAGCATCACCTGCACAGACCAGAAGGAGGAGCGTGTGATAGCAGC 530
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   486 CACTTTCCAGCAGCATCACCTGCACAGACCAGAAGGAGGAGCGTGTGATAGCAGC 545

Qy   531 ACACCGATGTTGTAACAAGAAATGGCATTGAGGAGGAGCGTCACAAACAGTAAAGTGTTTCCTG 590
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   546 ACACCGATGTTGCAACAAGAAATGGCATTGAGGAGGAGCGTCACAAACAGTAAAGTGTTTCCTG 605

Qy   591 TCTACCTGGAAAAGTGGCTGGGGACAACAAGAACCGGAGCCTTGCTCTTCGGCCTCCCAT 650
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   606 TCTACCTGGAAAAGTGGCTGGGGACAACAAGAACCGGAGCCTTGCTCTTCGGCCTCCCAT 665

Qy   651 AGTGATTGGGAAAATGGTGGTGTGAGAGATGGAGCCTTGCTCTAGAGGAGGAGAATGTAAGAC 710
          ||||||| ||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   666 AGTGATTTGGAAATGGGTGGTGTGAGAGATGGAGCCTTGCTCTAGAGGAGGAGAATGTAAGAC 725

Qy   711 ACTCCCTGACAATTCTGGATGGATGGTGTGCCTACAGGCAACACAAAATTAAGACCACGAGAAT 770
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   726 ACTCCCTGACAATTCTGGATGGATGGTGTGCCGACAGGCAACACAAAATTAAGACCACGAGAAT 785

Qy   771 TCACCCAAGAACCTAACACAGAAGCATTTGTTGGTAGTAAAGGAA 812
          |||||||||||||||||||||||||||||||||||||||||||||
Db   786 TCACCCAAGAACCTAACACAGAAGCATTTGTTGGTAGTAAAGGAA 827

RESULT 3
LOCUS   HSM803068   1936 bp   mRNA   linear   PRI 20-MAR-2002
DEFINITION Homo sapiens mRNA; cDNA DKFZp566B064 (from clone DKFZp566B064).
ACCESSION   AL713702
VERSION   AL713702.1 GI:19584419
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1936)
  AUTHORS   Wiemann,S., Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Gassenhuber,J. and
            Martinsried,
  TITLE   Direct Submission
  JOURNAL   Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152
            Martinsried, GERMANY

RESULT 2
LOCUS   AX086349   1936 bp   DNA   linear   PAT 09-MAR-2001
DEFINITION Sequence 301 from Patent WO0112659.

Left column

```
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by GBF (National Research Centre for Biotechnology Ltd.,
            Braunschweig/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp566B064) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES    Location/Qualifiers
     source 1..1936
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="DKFZp566B064"
            /tissue_type="kidney"
            /clone_lib="566 (synonym: hfkd2). Vector pAMP1; host
            X1-2blue; sites NotI + SalI"
            /dev_stage="fetal"
     gene   1..1936
            /gene="DKFZp566B064"
     CDS    400..801
            /gene="DKFZp566B064"
            /note="similarity to Homo sapiens clone 25003"
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="CAD28501.1"
            /db_xref="GI:19584420"
            /db_xref="SPTREMBL:Q8TCL8"
            /translation="MAMVSAMSWVLYLWISACAMLLCHGSLQHTFQQHHLHRPEGGTVC
            EVIAAHRCCNKNRIEERSQTVKCSCLPGKVAGTTRNRPSCVDASIVIWKWWCEMEPCL
            EGEECKTLPDNSGWMCATGNKIKTTRIHPRT"
     polyA_signal 1891..1896
            /gene="DKFZp566B064"
     polyA_site 1916
            /gene="DKFZp566B064"
ORIGIN

Query Match     53.3%;  Score 433.2;  DB 9;  Length 1936;
Best Local Similarity   96.1%;
Matches  444;  Conservative  0;  Mismatches  18;  Indels  0;  Gaps  0;

Qy  351 AGAGTACTCTCGTCTCTTGGCTGGATTTTTCAGAGAATGGCAATGGTCTCTCCGAGTCCTG 410
        |||| ||                     ||||||||||||||||||||||||| ||||||||
Db  366 AGGTAACCTTTCCTCGTCGAATGTTTCTTTAGAGAATGGCAATGGTCTCTCTGCGAGTCCTG 425

Qy  411 GGTCCTGTATTTGTGGGATAAAGTGCTTGGTGCGAAATGTCCCATGGTCTCTCTGCGAGCA 470
        |||||||||||||||||||||| ||||||| |||||  |||||||||||||||||||
Db  426 GGTCCTGTATTTGTGGGATAAGTGCTTGTGTGCAATGGTCTCTCTGCGAGTCCTTCAGCA 485

Qy  471 CACTTTCCAGCAGCATCACCTGCACAGACCAGGAAGGAGGAGGACGTGTGAAGTGAAGCAGC 530
        ||||||||||||||||||||||||||||  ||||||||||||  ||||  |||| ||||||
Db  486 CACTTTCCAGCAGCATCACCTGCACAGACCAGGAAGGAGGAGGACGTGTGAAGTGAAGCAGC 545

Qy  531 ACACCGGATGTTGTAACAAGAATCGCATTGAGGAGCGCGGTCAAACAGTAAAGTGTTCCTG 590
        |||||| ||||| |||||||||| |||||||||||||||||||||||||||||||||||||
Db  546 ACACCGATGTTGCACAAGAATCGCATTGAGGAGCGCGGTCAAACAGTAAAGTGTTCCTG 605

Qy  591 TCTACCTGGAAAAGTGGCTGGAACAACAAGAAAACCGGCCCTTCTTGGCGTCGATGCTCCAT 650
        ||||||| |||||||||||||||||||||| ||||| |||||||||||  |||| ||||||
Db  606 TCTACCTGGAAAAGTGGCTGGAACAACAAGAAAACCGGCCCTTCTTGGCGTCGATGCTCCAT 665

Qy  651 AGTGATTGGGAAATGTGGTGGTGTGAAATGTGAGAT 710
        ||||||||||||||| ||||||||||||||||||||
Db  666 AGTGATTGGGAAATGGTGGTGTGAAATGTGAGAT 725

Qy  711 ACTCCCTGACAATTCTGGATCGGATGTGGCGCAACAGCCACAAAAAATTAAGACCACAGAAT 770
        |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  726 ACTCCCTGACAATTCTGGATGTGGCGCAACAGCCACAAAAAATTAAGACCACAGAAT 785

Qy  771 TCACCCAAGAACCTAACAGAAGCATTTGTGGTAGTAAAGGA 812
        ||||||||||||||||||||||||||||||||||||||||||
```

Right column

```
Db  786 TCACCCAAGAACCTAACAGAAGCGATTTGTGGTAGTAAAGGAA 827

RESULT 4
LOCUS       AY325114          815 bp    mRNA    linear   PRI 23-JUL-2003
DEFINITION  Homo sapiens TAFA1 mRNA, complete cds.
ACCESSION   AY325114
VERSION     AY325114.1  GI:32967226
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 815)
  AUTHORS   Tang,Y.T., Emtage,P., Funk,W., Hu,T., Arterburn,M., Park,E. and
            Rupp,F.
  TITLE     TAFA: A Novel Secreted Family with Homology to CC-chemokines
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 815)
  AUTHORS   Tang,Y.T., Emtage,P., Funk,W., Hu,T., Arterburn,M., Park,E. and
            Rupp,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-2003) Biology Research, Nuvelo, 675 Almanor
            Avenue, Sunnyvale, CA 94085, USA
FEATURES    Location/Qualifiers
     source 1..815
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
     CDS    217..618
            /note="secreted protein similar to CC-chemokines"
            /codon_start=1
            /product="TAFA1"
            /protein_id="AAP92406.1"
            /db_xref="GI:32967227"
            /translation="MAMVSAMSWVLYLWISACAMLLCHGSLQHTFQQHHLHRPEGGTVC
            EVIAAHRCCNKNRIEERSQTVKCSCLPGKVAGTTRNRPSCVDASIVIGKWWCEMEPCL
            EGEECKTLPDNSGWMCATGNKIKTTRIHPRT"
ORIGIN

Query Match     53.2%;  Score 431.8;  DB 9;  Length 815;
Best Local Similarity   99.5%;
Matches  433;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy  378 TCAGAGAATGGCAATGGTCTCTCCGAGTCCTGGGTCCTGTATTTGTGGGATAAGTGCTTG 437
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  210 TTATAGAATGGCAATGGTCTCTCTGCGAGTCCTGGGTCCTGTATTTGTGGGATAAGTGCTTG 269

Qy  438 TGCAATGCTACTCTGCCATGGATCCCTTCAGCACACTTTCCAGCAGCATCACCTGCACAG 497
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  270 TGCAATGCTACTCTGCCATGGATCCCTTCAGCACACTTTCCAGCAGCATCACCTGCACAG 329

Qy  498 ACCAGGAAGGAGGAGGACGTGTGAAGTGAAGCAGCACACCGATGTTGTAACAAGAATCGCAT 557
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  330 ACCAGGAAGGAGGAGGACGTGTGAAGTGAAGCAGCACACCGATGTTGTAACAAGAATCGCAT 389

Qy  558 TGAGGAGCGCGGTCACAAAACAGTAAAGTGTTCCTGTCTACCTGGAAAAGTGGCTGGAACAAC 617
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  390 TGAGGAGCGCGGTCACAAAACAGTAAAGTGTTCCTGTCTACCTGGAAAAGTGGCTGGAACAAC 449

Qy  618 AAGAAAACCGGCCCTTCTTGGGTCGATGCTCCATAGTGATTGGGAAATGGTGGTGTGAAGAT 677
        |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  450 AAGAAAACCGGCCCTTCTTGGTCGATGCTCCATAGTGATTGGGAAATGGTGGTGTGAAGAT 509

Qy  678 GGAGCCTTGCCTAGAAGGAGGAGAAGAAATGTAAGACACTCCCCTGACAATTCTGGATGGGATCTG 737
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  510 GGAGCCTTGCCTAGAAGGAGGAGAAGAAATGTAAGACACTCCCCTGACAATTCTGGATGGGATGTG 569

Qy  738 CGCAACAGGCAACAAAAAATTAAGACCACAGAGAATCAACCCAAGAACCTAACAGAAGCATTT 797
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  570 CGCAACAGGCAACAAAAAATTAAGACCACAGAGAATCACCCAAGAACCTAACAGAAGCATTT 629
```

```
Qy      798 GTGGTAGTAAAGGAA 812
            |||||||||||||||
Db      630 GTGGTAGTAAAGGAA 644
```

```
RESULT 5
LOCUS       AC092036           137703 bp    DNA     linear   PRI 13-SEP-2001
DEFINITION  Homo sapiens chromosome 3 clone RP11-79C12, complete sequence.
ACCESSION   AC092036
VERSION     AC092036.3  GI:15594336
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 137703)
AUTHORS     Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C.,
            Clendenning,J. and Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 137703)
AUTHORS     Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J., Ivey,R.G. and
            Haugen,B.D.
TITLE       Direct Submission
JOURNAL     Submitted (16-JUN-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 137703)
AUTHORS     Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C.,
            Clendenning,J. and Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (13-SEP-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Sep 13, 2001 this sequence version replaced gi:15138904.
            Center: University of Washington Genome Center
            Center Code: UWGC
            Web site: http://www.genome.washington.edu
            Contact: uwgchtgs@u.washington.edu
            Drafting Center: BCM
            -------------- Project Information
            Center project name: chr-3
            Center clone name: RP11-79C12 (bc0194)
            -------------- Summary Statistics
            Sequencing vector: M13; L08821; 60% of reads
            Sequencing vector: plasmid; L08752; 40% of reads
            Chemistry: Dye-primer BigDye; 36% of reads
            Chemistry: Dye-terminator ET; 36% of reads
            Chemistry: Dye-terminator Big Dye; 28% of reads
            Assembly Program: Phrap; version 0.990319
            Consensus quality: 139753 bases at least Q40
            Consensus quality: 140133 bases at least Q30
            Consensus quality: 140304 bases at least Q20
            Insert size: 137703; sum-of-contigs
            Quality coverage: 9.6x in Q20 bases; sum-of-contigs
            --------------
            Overlapping Sequences:
            5': RP11-81N13 (UWGC:bc0203) AC016935
            3': RP11-364N10 (UWGC:bc0385) AC021437
            --------------
            Sequence Quality Assessment:
            This entry has been annotated with sequence quality
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality zero.
            Quality levels above 40 are expected to have less than
            1 error in 10,000 bp.
            Base-by-base quality values are not generally visible from the
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.
            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            Sequence Validation:
            This sequence has been validated by Multiple Complete Digest
            fingerprinting. Comparison of the experimentally derived digest
            fragments with sequence-predicted fragments is given below.
            The electronically-digested sequence consists of both insert and
            vector, in order to accurately represent the entire circular BAC.
            Small fragments below a variable cutoff (approximately 400-800 bp)
            are not resolved in the fingerprint and hence do not appear
            in the table. There are no significant remaining discrepancies
            between the experimental and predicted values. Uniquely ordered
            fragments are separated by dashed lines.
```

BglII		EcoRI		HindIII	
SeqDerMap	FngrPrint	SeqDerMap	FngrPrint	SeqDerMap	FngrPrint
2273	2272	8696	8904	2190	2099
2067	2102	6	<800	6382	6589
6314	6524	159	<800	512	<800
813	778	1461	1486	449	<800
9138	8978	839	848	5669	5435
4117	4135	1014	994	150	<800
5429	5600	458	<800	680	<800
3261	3286	1505	1486	457	<800
11041	10746	968	994	3181	3182
9336	9295	55	<800	3742	3733
92	<800	2448	2471	1207	1241
6268	6269	1679	1736	2016	1899
5523	5600	6963	6941	56	<800
17081	17253	8613	8629	2083	2099
509	<800	4027	3975	3015	3182
768	778	6850	6941	3204	3306
1815	1827	2003	2026	5520	5435
2712	2727	9934	9883	536	<800
5619	5600	1410	1486	2744	2830
1410	1393	1784	1736	460	<800
6945	7000	1225	1249	3609	3588
1947	1947	1719	1736	2707	2830
286	<800	698	<800	10577	10789
2123	2102	3448	3300	892	879
501	<800	3219	3300	5440	5435
5108	5108	1739	1736	6194	6078

RESULT 6
AC138390

```
LOCUS       AC138390              158080 bp    DNA     linear   PRI 22-JAN-2003
DEFINITION  Homo sapiens chromosome 3 clone RP11-81N13, complete sequence.
ACCESSION   AC138390 AC016935
VERSION     AC138390.2  GI:27819541
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158080)
AUTHORS     Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
            Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 158080)
AUTHORS     Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (28-DEC-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 158080)
AUTHORS     Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
            Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (22-JAN-2003) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Jan 22, 2003 this sequence version replaced gi:27413178.
            -------------------- Genome Center
            Center: University of Washington Genome Center
            Center Code: UWGC
            Web site: http://www.genome.washington.edu
            Contact: uwgchtgs@u.washington.edu
            Drafting Center: BCM
            -------------------- Project Information
            Center project name: chr-3
            Center clone name: RP11-81N13 (bc0203)
            -------------------- Summary Statistics
            Sequencing vector: plasmid; 46% of reads
            Sequencing vector: M13; L08821; 54% of reads
            Chemistry: Dye-terminator Big Dye; 99% of reads
            Chemistry: Dye-primer BODIpy; 1% of reads
            Assembly Program: Phrap; version 0.990319
            Chemistry quality: 158072 bases at least Q40
            Consensus quality: 158080 bases at least Q40
            Consensus quality: 158080 bases at least Q30
            Consensus quality: 158080 bases at least Q20
            Insert size: 158080; sum-of-contigs
            Quality coverage: 11.0x in Q20 bases; sum-of-contigs
            --------------------
```

Query Match   45.4%; Score 368.8; DB 9; Length 137703;
Best Local Similarity   98.7%;
Matches 383; Conservative 0; Mismatches 2; Indels 3; Gaps 1;
Pred. No. 4.2e-87;

```
Qy   121 AGGTTGAAAATGACAAGGTTTCCACCCCTCCAAAACCTTGCCTTCCTTTTCTGACAATACAGTC 180
         ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db 10111 AGGTTGAAAAATGACAAGGTTTCCACCCCTCCAAAACCTTGGCTCCTTTTCTGACAATACAGTC 10052

Qy   181 TGAATGAACCGATGTCTTTT------ACTTGTGAAATAGGATCGGAAGAGAGAGTAACA--- 237
         ||||||||||||||||||||      ||||||||||||||||||||||||||||||||||
Db 10051 TGAATGAACCGATGTCTTTTTTTTTTACTTGTGAAATAGGATCGGAAGAGAGAGTAACATTT 9992

Qy   238 TTTTTTTTTTTTTAATCCTGATAAAAGAAGATTGTTGGGAAGGCTCTTTGAAAAAAAATTTTAAA 297
         ||||||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  9991 TTTTTTTT--TAATCCTGATAAAAGAAGATTGTTGGGAAGGCTCTTTGAAAAAAAATTTTAAA 9932

Qy   298 TTGTGGCACAGATGGATTTTAAAAGTGTTAGATCTTTCCAATGAAACACTAAATAGAGTAC 357
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  9931 TTGTGGCACAGATGGATTTTAAAAGTGTTAGATCTTTCCAATGAAACACTAAATAGAGTAC 9872

Qy   358 TCTGCTCTTGGCTGATTTTTCAGAGA 385
         ||||||||||||||||||||||||||
Db  9871 TCTGCTCTTGGCTGATTTTTCAGAGTAA 9844
```

```
3326 ------ 3286      4504 ------ 4495      1504 ------ 1474      2611 ------ 2646
5600 ------ 5600       715 ------ <800       873 ------  879       419 ------ <800
 116 ------ <800      1791 ------ 1736      1299 ------ 1241      3343 ------ 3300
2897 ------ 2948      3659 ------ 3666      3318 ------ 3306       443 ------ <800
5688 ------ 5600       893 ------  826      2913 ------ 3020      9705 ------ 9605
 771 ------  778      2787 ------ 2796       423 ------ <800      1548 ------ 1486
2951 ------ 2948      1973 ------ 2026       113 ------ <800      5322 ------ 5250
2714 ------ 2727      7554 ------ 7421       117 ------ <800       821 ------  826
  91 ------ <800      4961 ------ 4991       426 ------ <800
4960 ------ 4859       281 ------ <800      1913 ------ 1899
4204 ------ 4135      5348 ------ 5250      5403 ------ 5435
 588 ------ <800      7332 ------ 7421        42 ------ <800
```

```
2154 ------ 2099      1691 ------ 1736      5781 ------ 5817
  12 ------ <800                             9136 ------ 9068
4489 ------ 4371       556 ------ <800
2882 ------ 2830
 677 ------ <800       885 ------  879
 217 ------ <800       204 ------ <800
4592 ------ 4596       799 ------  879
 464 ------ <800       585 ------ <800
                      2100 ------ 2099
                      4115 ------ 4147
                      4315 ------ 4371
                      1823 ------ 1796
```

```
Qy   1    CACTGGAGTGGGGATGGGTCCATGGACGACTATATAAACTGATTCTTCATCAGGAAACTGCACA 60
          |||||||||||||||||||||||| ||  |||||||||||||||| ||||||||||||||||
Db 10231  CACTGGAGTGGGGATGGGTCCATCGGCAACTATAAAACTGATTCTCATCAGGAAACTGCACA 10172

Qy  61    TTATCTCCCCATCACTTCAAAGGTTCTTCGTCAGGAGGAGGTGACGCCAGGAGGAGATGATTTAA 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 10171  TTATCTCCCCATCACTTCAAAGGTTCTTCGTCAGGAGGAGGTGACGCCAGGAGGAGATGATTTAA 10112
```

Overlapping Sequences:
5': RP11-79C12 (UWGC:bc0194) AC092036, 134230-bp overlap
3': RP11-89A12 (UWGC:bc0207) AC092416, 45454-bp overlap

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

| EcoRI | | HindIII | | BglII | |
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
8696	8768	7698	7651	5136	5080
6	<800	6382	6470	2067	2084
458	<800	512	<800	12792	12692
1505	1524	449	<800	4117	4117
968	946	2196	2138	5429	5542
55	<800	150	<800	3261	3294
2448	2480	680	<800	11041	10803
1679	1712	457	<800	9336	9287
6963	6907	3181	3188	92	<800
8613	8768	3742	3737	6268	6223
4027	3995	1207	1229	5523	5542
6850	6907	2016	1912	17081	17289
2003	2014	56	<800	509	<800
9934	9693	2083	2138	768	779
1410	1413	3015	3023	1815	1794
1784	1712	3204	3188	2712	2734
1225	1225	5520	5367	5619	5542
1719	1712	536	<800	1410	1367
698	<800	2744	2824	6945	6947
3448	3361	460	<800	1947	1919
3219	3609	286	<800		
1739	1712	2707	2824	2123	2084
4504	4516	10577	10818	501	<800
715	892	892	893	5108	5080
1791	1712	5440	5367	3326	3294
3659	3686	6194	6047	5600	5542
4961	4996	2913	2824	2951	2936
281	<800	423	<800	2714	2734
7554	7429	3318	3316	771	779
1973	2014	1299	1229	5688	5542
2787	2820	873	893	2897	2936
893	946	1504	1447	116	<800
5348	5258	113	<800	91	<800
7332	7429	117	<800	4960	4824
5781	5807	426	<800	4204	4117
1691	1712	1913	1912	588	<800
2611	2663	5403	5367	7730	7706
419	<800	42	<800	2853	2936
3343	3361	9136	9077	1229	1203
443	<800	556	<800	6500	6512
9705	9693	2154	2138	1218	1203
1548	1524	12	<800	1102	1093
5322	5258	4489	4377	81	<800
1425	1413	677	<800	274	<800
818	828	2882	2824		
6636	6599	217	<800		
714	<800	4592	4557		
1243	1225	799	<800		
6726	6907	464	<800		
5258					
1524					
9693					
3504	3361	204	<800		
3363	3361	885	893		
242	<800				
585	<800	2100	2138		
1225	1225	4115	4117		

RESULT 7

```
                                -------
                                   4315      4377
                                -------
                                   1823      1776
                                -------
```

Query Match   45.4%; Score 368.8; DB 9; Length 158080;
Best Local Similarity 98.7%; Pred. No. 4.2e-87;
Matches 383; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

```
Qy       1 CACTGGAGTGGGGGATGGTCCATCGGCCAACTATAAACTGATTCTCATCAGGAAACTGCACA   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124000 CACTGGAGTGGGGGATGGTCCATCGGCCAACTATAAACTGATTCTCATCAGGAAACTGCACA 124059

Qy      61 TTATCTCCCCATCACTTCAAAAGTTCTCGTCAGCAGAGAGGTGACGCCCAGGAGATGATTTAA  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124060 TTATCTCCCCATCACTTCAAAAGTTCTCGTCAGCAGAGAGGTGACGCCCAGGAGATGATTTAA 124119

Qy     121 AGGTGAAAAATGACAAAGGTTTCCACCCTCAAACCTTGGCTCCTTTCTGACAATACAGTC  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124120 AGGTGAAAAATGACAAAGGTTTCCACCCTCAAACCTTGGCTCCTTTCTGACAATACAGTC 124179

Qy     181 TGAATGAACCCGAGTGTCTTTTTTTTACTGTGGAAATAGGATGGGAGAGAGTAACA--- 237
           |||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db  124180 TGAATGAACCCGAGTGTCTTTTTTTTACTGTGGAAATAGGATGGGAGAGAGTAACATTT 124239

Qy     238 TTTTTTTTTTAATCCTGATAAAGAAGAAGATTGTTGGGAAGCTCTTTGAAAAAAATTTTAAA  297
           ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db  124240 TTTTTTTTTTAATCCTGATAAAGAAGAAGATTGTTGGGAAGCTCTTTGAAAAAAATTTTAAA 124299

Qy     298 TTGTGGCACAGATGGATTTTAAAAAGTTGTTAGATCTTTCCAATGAACACTAATAGAGTAC  357
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124300 TTGTGGCACAGATGGATTTTAAAAAGTTGTTAGATCTTTCCAATGAACACTAATAGAGTAC 124359

Qy     358 TCTGCCTCTTGGCTCGGATTTTTTCAGGTAA  385
           |||||||||||||||||||||||||||||||
Db  124360 TCTGCCTCTTGGCTCGGATTTTTTCAGGTAA 124387
```

LOCUS   AC139766   187748 bp   DNA   linear   PRI 20-MAR-2003
DEFINITION Homo sapiens 3 BAC RP11-68P9 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC139766
VERSION   AC139766.15 GI:29126299
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 187748)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.P., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Deferich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Barnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,R., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,M., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Morris,R.T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Teifroo,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 187748)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (13-FEB-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 187748)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (19-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4 (bases 1 to 187748)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (20-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Mar 20, 2003 this sequence version replaced gi:29123832.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
              STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
              Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
              Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.

Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht
ml.

```
FEATURES             Location/Qualifiers
     source          1..187748
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-688P9"
     repeat_region   52..82
                     /rpt_family="(TA)n"
     repeat_region   complement(83..379)
                     /rpt_family="AluSq"
     repeat_region   467..691
                     /rpt_family="L2"
     repeat_region   1206..1312
                     /rpt_family="Tigger6"
     repeat_region   1995..2012
                     /rpt_family="AT_rich"
     repeat_region   2015..2744
                     /rpt_family="L1MA3"
     repeat_region   2745..2873
                     /rpt_family="MER85"
     repeat_region   2874..4418
                     /rpt_family="L1MA3"
     repeat_region   4422..5898
                     /rpt_family="L1PA8"
     repeat_region   5904..6369
                     /rpt_family="L1MA3"
     repeat_region   6372..6397
                     /rpt_family="AT_rich"
     repeat_region   7091..7518
                     /rpt_family="L1ME3B"
     repeat_region   7575..7597
                     /rpt_family="(TC)n"
     repeat_region   7598..7702
                     /rpt_family="(TA)n"
     repeat_region   7892..7938
                     /rpt_family="L1ME3B"
     repeat_region   8102..8124
                     /rpt_family="AT_rich"
     repeat_region   complement(8554..8981)
                     /rpt_family="(CAAAA)n"
     repeat_region   complement(8992..9346)
                     /rpt_family="MLT1C"
     repeat_region   9385..9900
                     /rpt_family="THE1A"
     repeat_region   9906..10003
                     /rpt_family="L1ME3B"
     repeat_region   10148..10173
                     /rpt_family="AT_rich"
     repeat_region   10255..10489
                     /rpt_family="MIR"
     repeat_region   10820..10853
                     /rpt_family="AT_rich"
     repeat_region   complement(11238..11419)
                     /rpt_family="MIRb"
     repeat_region   12049..12270
                     /rpt_family="MIR"
     repeat_region   12516..12616
                     /rpt_family="MIR"
     repeat_region   13267..13439
                     /rpt_family="L2"
     repeat_region   13479..13509
                     /rpt_family="L2"
     repeat_region   13539..13620
                     /rpt_family="(CAA)n"
     repeat_region   13621..13989
                     /rpt_family="MER20"
     repeat_region   complement(14577..14883)
                     /rpt_family="L1MA8"
     repeat_region   15436..15578
                     /rpt_family="AluSx"
     repeat_region   15580..16677
                     /rpt_family="MER33"
     repeat_region   complement(16686..16990)
                     /rpt_family="L1MA8"
     repeat_region   17006..17214
                     /rpt_family="L2"
     repeat_region   17634..17724
                     /rpt_family="MER33"
     repeat_region   17841..17993
                     /rpt_family="MIR"
     repeat_region   17996..18218
                     /rpt_family="L1MA4A"
     repeat_region   complement(18372..18456)
                     /rpt_family="MIR3"
     repeat_region   complement(18508..18736)
                     /rpt_family="MIRb"
     repeat_region   19223..19251
                     /rpt_family="AT_rich"
     repeat_region   complement(19617..19914)
```

Query Match   45.4%; Score 368.8; DB 9; Length 187748;
Best Local Similarity 98.7%; Pred. No. 4.2e-87;
Matches 383; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

```
Qy        1 CACTGGGAGTGGGGGGATGGTGCCATCGGGCAACTATAAAACTGATTCTCATCAGGAAACTCGCACA 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    48213 CACTGGGAGTGGGGGGATGGTGCCATCGGGCAACTATAAAACTGATTCTCATCAGGAAACTCGCACA 48272

Qy       61 TTATCTCCCCATCACTTCAAAGGTCTCGTCAGGCAGAGGGTGACGCCAGGAGGATGAGTTAA 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    48273 TTATCTCCCCATCACTTCAAAGGTCTCGTCAGGCAGAGGGTGACGCCAGGAGGATGAGTTAA 48332

Qy      121 AGGTGAAAATGACAAGGTTTCCACCCCTCAAACCTTGGCTCCTTTCTTTGACAATACAGTC 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    48333 AGGTGAAAATGACAAGGTTTCCACCCCTCAAACCTTGGCTCCTTTTTCTGACAATACAGTC 48392

Qy      181 TGAATGAACCCGATGTCTTTTTTTTTACTGTGGAAATAGGATGGAAGGAGTAACA--- 237
            |||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    48393 TGAATGAACCCGATGTCTTTTTTTTTTACTGTGGAAATAGGATGGAAGGAGTAACAATT 48452

Qy      238 TTTTTTTTTTTTTTAATCCTGATAAAGCAAGAGATTGTTGGGGAAGCTCTTTTGAAAAAAAATTTTAAA 297
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    48453 TTTTTTTTTTTTAATCCTGATAAAGCAAGAGATTGTTGGGGAAGCTCTTTTGAAAAAAAATTTTAAA 48512

Qy      298 TTGTGGGCACAGATGGATTTTTTAAAAAGTCTTTAGATCTTTCCAATGAAACACTAATTAGAGTAC 357
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    48513 TTGTGGGCACAGATGGATTTTTTAAAAAGTCTTTAGATCTTTCCAATGAAACACTAATTAGAGTAC 48572

Qy      358 TCTGCTCTTGGCCTGGATTTTTCAGGAGAA 385
            ||||||||||||||||||||||||||||||
Db    48573 TCTGCTCTTGGCCTGGATTTTTCAGGAGTAA 48600
```

RESULT 8
AC116080/c
AC116080

LOCUS   AC116080   247515 bp   DNA   linear   HTG 11-OCT-2002
DEFINITION Rattus norvegicus clone CH230-40C18, *** SEQUENCING IN PROGRESS
            ***.
ACCESSION   AC116080
VERSION   AC116080.4 GI:23664937
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE   Rattus norvegicus
  ORGANISM Rattus norvegicus (Norway rat)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 247515)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,

REFERENCE
AUTHORS   Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
           Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
           Bryant,N., Buhay,C., Burch,P., Burrell,A., Calderon,E.,
           Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
           Chavez,D., Chen,G., Chen,Y., Chen,Z., Chu,J.,
           Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,J.,
           Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
           Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
           Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
           Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
           Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
           Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
           Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
           Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
           Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
           Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
           Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
           Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,H., Jolivet,A.,
           Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
           Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
           Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
           Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
           Maheshwari,M., Mahindarne,M., Mahmoud,M., Malloy,K., Mangum,A.,
           Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
           Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
           Miosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
           Morgan,M., Morris,M., Morris,M., Munidasa,M., Murphy,M., Nair,L.,
           Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
           Nwokelemeh,O., Okwuonu,G., Olairunsagoon,A., Pal,S., Parks,K.,
           Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
           Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.L.,
           Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
           Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
           Rives,C., Rogkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.,
           Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
           Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
           Sneed,J., Sodergren,E., Song,X.-Z., Sorelle,M., Sosa,J.,
           Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
           Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
           Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
           Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
           Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
           Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,C., Yoon,V.,
           Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
           Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
           Weinstock,G. and Gibbs,R.A.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 247515)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (24-MAR-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 247515)
AUTHORS   Rat Genome Sequencing Consortium.
TITLE   Direct Submission
JOURNAL   Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Oct 10, 2002 this sequence version replaced gi:21736969.
          The sequence in this assembly is a combination of BAC based reads
          and whole genome shotgun sequencing reads assembled using Atlas
          (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
          in the feature table below represents a scaffold in the Atlas
          assembly (a 'contig-scaffold'). Within each contig-scaffold,
          individual sequence contigs are ordered and oriented, and separated
          by sized gaps filled with Ns to the estimated size. The sequence
          may extend beyond the ends of the clone and there may be sequence
          within a contig-scaffold that consist entirely of whole
          genome shotgun sequence reads. Both end sequences and whole genome
          shotgun sequence only contigs will be indicated in the feature
          table.

----------------------- Genome Center
          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu/
          Contact: hgsc-help@bcm.tmc.edu
----------------------- Project Information
          Center project name: GNWA
          Center clone name: CH230-40C18
----------------------- Summary Statistics
          Assembly program: Phrap; version 0.990329
          Consensus quality: 229334 bases at least Q40
          Consensus quality: 219595 bases at least Q30
          Consensus quality: 232872 bases at least Q20
          Consensus quality: 236904; sum-of-contigs estimation
          Estimated insert size: 236904; sum-of-contigs estimation
          Quality coverage: 6x in Q20 bases;
----
*   NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank draft data.html).
*   NOTE: This is a 'working draft' sequence. It currently
*   consists of 1 contigs. Gaps between the contigs
*   are represented as runs of N. The order of the pieces
*   is believed to be correct as given, however the sizes
*   of the gaps between them are based on estimates that have
*   provided by the submitter.
*   This sequence will be replaced
*   by the finished sequence as soon as it is available and
*   the accession number will be preserved.
*
*   1 247515: contig of 247515 bp in length.
FEATURES   Location/Qualifiers
     source   1..247515
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-40C18"
     misc_feature   1..1162
                     /note="wgs end_extension
                     clone_end:T7"
     misc_feature   2343..4221
                     /note="wgs end_extension
                     clone_end:T7"
     misc_feature   5130..6933
                     /note="wgs end_extension
                     clone_end:T7"
     misc_feature   8859..9540
                     /note="clone_boundary
                     clone_end:T7"
                     site:EcoRI
                     end_sequence:BH315138"
     misc_feature   181526..183635
                     /note="wgs_contig"
     misc_feature   243698..245700
                     /note="wgs_contig"
     misc_feature   245971..247515
                     /note="wgs_contig"

ORIGIN
     Query Match
     Best Local Similarity   82.1%;   Pred. No. 3.2e-46;
     Matches 325;   Conservative   0;   Mismatches   53;   Indels   18;   Gaps   6;

Qy   1 CACTGGAGTGGGGATGGTCCATCGGCAACTATAAACTGATTCTAATCAGGAAACTGCACA 60
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db 141636 CACTGGAGTAGGGATGGTCTGTCCGTGGCTTATAAACTGCTTCTTCAGGAAACT--ACA 141579

Qy   61 TTATCTCCCCATCACTTCAAAAGGTCTCGTCAGGCAGAGGTGACCGCCAGGAGATGATTAA 120
          || |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db 141578 TTAGCTCCACCATCACTTCAAAAGGTCTCGTCAGGCAGAGGTGACCGCCAGGGGCATGATTTAA 141519

Qy   121 AGGTGAAAATGACAAGGGTTTCCACCCCCCCAAAACCTTGGCTCCCTTTTCTGACAATACAGTC 180
          ||| ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db 141518 AGGTGAACATGACAAGGTTTCCACCCCTCAAAACCTTGGTTCCTTTTCTGACAATACAG-- 141461

Qy   181 TGAATGAACCCGATGTCTTTTTTTTTTTTTTTTACTGTGTGGAAAAATAGGATCGGAAGAGAGTAACATTT 240

```
Db      141460  TGAACAGAACCCAGCGTC---TTCACAACTGTGGAAACAGGACTGGAAGAGGGAAGCGTGC 141404
                |||  |||||||                |||     |||||||||||||  ||||||||||||  ||
Qy      241     TTTTTTT-------AATCCTGATAAAGAA--GATTGTTGGGAAGCATCTCTTTGAAAAAAAAA 290

Qy      291     TTTTA-AATTGTGGCACAGATTGGATTTTAAAAGTGTTAGATCTTTCCAATGAACACTAA 349
                ||||||| ||||||||||||||  ||||| ||||||||||||||||||  |||||||
Db      141403  TTTTTTTTTTCCTTCAATTCCTAATAAAGAACGTGCTGGGAAGCATTCTCTTTGAAAAAAAA 141344

Qy      350     TAGAGTACTCTGCTCCTTGGCCTGGAATTTTTCAGAGAA 385
                ||||| ||||||||||||||||||||||||||||||||||||| ||
Db      141283  GAGAGTGTCCTTGCCTTGGCCTGGGAATTTTCAGGTAA 141248

Db      141343  TCTCAGAACTGTGGGCACAGAGGAGTTTAAAAAGTTAGCTCTTTCCAATGAACACTAG 141284
```

```
RESULT 9
LOCUS       AC117234                193027 bp    DNA     linear   ROD 25-NOV-2003
DEFINITION  Mus musculus BAC clone RP23-410J2 from chromosome 6, complete
            sequence.
ACCESSION   AC117234
VERSION     AC117234.4  GI:34495048
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 193027)
  AUTHORS   Swearengen-Shahid,S., Kozlowicz,A. and Schatzkamer,K.
  TITLE     The sequence of Mus musculus BAC clone RP23-410J2
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 193027)
  AUTHORS   Wilson,R.
  TITLE     Sequencing of Mus musculus
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 193027)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 193027)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 193027)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   6  (bases 1 to 193027)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Sep 6, 2003 this sequence version replaced gi:21735140.
            -----------  Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -----------  Summary Statistics
                Center project name: M_BA0410J02
            -----------
            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa
            and Minako Tateno in the laboratory of Pieter de Jong
            (http://www.chori.org) from female C57BL/6J mouse kidney and/or
            brain genomic DNA.  The clone and detailed information can be
            obtained from Research Genetics, Inc. (http://www.resgen.com) or
            Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1..193027
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="6"
                     /map="6"
                     /clone="RP23-410J2"
                     /clone_lib="RPCI-23"
     repeat_region   1..128
                     /rpt_family="L1"
     repeat_region   129..197
                     /rpt_family="B4"
     repeat_region   528..737
                     /rpt_family="B4"
     repeat_region   959..1031
                     /rpt_family="MER2_type"
     repeat_region   1032..1075
                     /rpt_family="ID"
     repeat_region   1075..1031
                     /rpt_family="B4"
     repeat_region   3251..3605
                     /rpt_family="L1"
     repeat_region   3640..3834
                     /rpt_family="L1"
     repeat_region   3917..3988
                     /rpt_family="MER1_type"
     repeat_region   4026..4243
                     /rpt_family="L1"
     repeat_region   4708..4969
                     /rpt_family="RMER15"
     repeat_region   4903..5013
                     /rpt_family="B4"
     repeat_region   5933..6355
                     /rpt_family="MaLR"
     repeat_region   7471..7535
                     /rpt_family="MIR"
     repeat_region   8988..9064
                     /rpt_family="ERVK"
     repeat_region   9260..9476
                     /rpt_family="MaLR"
     repeat_region   9600..9719
                     /rpt_family="L2"
     repeat_region   11657..11695
                     /rpt_family="L1"
     repeat_region   11697..12125
                     /rpt_family="L1"
     repeat_region   12124..12182
                     /rpt_family="L1"
     repeat_region   12185..12849
                     /rpt_family="L1"
     repeat_region   12860..12895
                     /rpt_family="L1"
```

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

```
repeat_region  14641..14724   /rpt_family="L1"
repeat_region  14712..15744   /rpt_family="L1"
repeat_region  15767..15806   /rpt_family="L1"
repeat_region  16793..16958   /rpt_family="L2"
repeat_region  17034..17198   /rpt_family="MIR"
repeat_region  18190..18313   /rpt_family="MIR"
repeat_region  20721..21086   /rpt_family="L2"
repeat_region  21368..21506   /rpt_family="L2"
repeat_region  22720..22980   /rpt_family="L1"
repeat_region  24309..24434   /rpt_family="B4"
repeat_region  2497?..25060   /rpt_family="B4"
repeat_region  25861..25973   /rpt_family="MIR"
repeat_region  29872..30005   /rpt_family="Alu"
repeat_region  30021..30096   /rpt_family="B4"
repeat_region  31318..31684   /rpt_family="ID"
repeat_region  31935..32071   /rpt_family="MaLR"
repeat_region  32532..32617   /rpt_family="L2"
repeat_region  32871..33150   /rpt_family="MER1_type"
repeat_region  33151..33421   /rpt_family="MaLR"
repeat_region  34099..34366   /rpt_family="B4"
repeat_region  34368..34521   /rpt_family="L1"
repeat_region  35500..35548   /rpt_family="L1"
repeat_region  37001..37148   /rpt_family="ERV1"
repeat_region  37169..37439   /rpt_family="Alu"
repeat_region  37403..37493   /rpt_family="B4"
repeat_region  37515..38019   /rpt_family="B2"
repeat_region  38034..38279   /rpt_family="L1"
repeat_region  38280..38425   /rpt_family="ERVK"
repeat_region  38426..38502   /rpt_family="Alu"
repeat_region  38503..38609   /rpt_family="L1"
repeat_region  40100..40501   /rpt_family="ERVK"
repeat_region  41214..41406   /rpt_family="MaLR"
repeat_region  41548..41737   /rpt_family="MaLR"
repeat_region  42486..42881   /rpt_family="ERVK"
repeat_region  43220..43399   /rpt_family="ERVK"
repeat_region  44489..44558   /rpt_family="MIR"
```

```
Qy    2  ACTGGAGTGGGGAATGGTTCCATCGGCCAACTATATAAAACTGATTCATCAGAAACTGCACAT   61
         ||||||   |||||||||| |  || || ||||| |||||| ||||||||||||||| ||||
Db 101311 ACTGGAGTAGGGATGGTCGTCCGTGGGCTATAAACTGCTTCATCAGAAACT--ACAT 101254

Qy   62  TATCTCCCCATCACTTCAAAGGTCTTCGTCAGGCAGGAGGTGACGGCCAGGAGGATGAGTTAAA  121
         |||||||| |||||||||||||||| |||||| |||||||||| ||||||||| |||||| |||
Db 101253 TAGCTCACCCATCACTTCAAAGGTCTTCGTCAGGCAGGAGGTGACGGCCAGGAGGATGAGTTAAA 101194

Qy  122  GGTGAAAAATGACAAGGTTTTCCACCCCTCAAACCTTGGCTCCTTTTTCTGACAAATACAGTCT  181
         ||||||||||||||||||||||||||||||||| |||||||||||||  |||||||| ||||
Db 101193 GGTGAACATGACAAGGTTTCCCACCCCTCAAACCTTGCTTCCTTTTTCTGACAAATACAG--T  101136

Qy  182  GAAATGAAACCCGATGTCTTTTTTTTTTTTTACTGTGGAAATAGGATCGGAAGAGAGTAAC----  236
         ||   |||||||||  |||||||||||||   ||||||||   ||||| |||||||| ||
Db 101135 GAACGAGCCCAAACGTC---TTCACAACTGTGGAAACGGGACTGGAAGAAAAAGCTTGCC 101079

Qy  237  ------------------AATTTTTTTTTTATCCTGATAAAGAAGATTGTTG  270
                           |||||||||| |||||||| ||||||||||||
Db 101078 TTTTTTTTCTTTTTTTTTCTTTTTTTTTTTTTCAATTCTTAATAAAGATAATTGCCTGGG 101019

Qy  271  GGAAGCTGCTTTGAAAAAAAATTTTAAATTGTGGCACAGATGGGATTTTAAAAGTGTTAGA  330
         ||||||||  ||||||||||||||  ||||||||||||||||||||||||||||||| |||
Db 101018 AAGCAATTCTCTTTGAAAAAATCTCAGAACGTGGGCACAGATGGGATTTTAAAAAGTGTTAGC 100959

Qy  331  TCTTTCCAATGAACACTAATAGAGTACTCTGCTCTTGGCGTGGAGTTTTTCAGGAGAA  385
         || || ||||||||||||||   | || || |||  || ||||||||||||||||
Db 100958 TCTTTCCAATGAGCGACTACTAGGAGGGTTCCCTGCTTCTTGGCTGGGATTTTTCAGGTAA 100904
```

```
RESULT 10
AX896885
LOCUS        AX896885         261 bp    DNA         linear    PAT 18-DEC-2003
DEFINITION   Sequence 12748 from Patent EP1033401.
ACCESSION    AX896885
VERSION      AX896885.1    GI:40051769
KEYWORDS     .
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1
  AUTHORS    Dumas Milne Edwards,J.B., Duclert,A. and Giordano,J.Y.
  TITLE      Expressed sequence tags and encoded human proteins
  JOURNAL    Patent: EP 1033401-A 12748 06-SEP-2000;
             Genset (FR)
FEATURES             Location/Qualifiers
     source          1..261
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN
```

```
Query Match
Best Local Similarity 98.9%;   Score 187;  Pred. No. 9.4e-39;  Indels 0;  Gaps 0;
Matches 187; Conservative 0; Mismatches 2;

Qy    1  CACTGGAGTGGGGAATGGTTCCATCGGCCAACTATATAAAACTGATTCATCAGAAACTGCACA   60
         |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73  CACTGGAGTANNGGGGAATGGTCCATCGGCCAACTATAAAACTGATTCATCAGAAACTGCACA 132

Qy   61  TTATCTCCCCATCACTTCAAAGGTCTTCGTCAGGCAGGAGGTGACGGCCAGGAGGATGAGTTTAA  120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133  TTATCTCCCCATCACTTCAAAGGTCTTCGTCAGGCAGGAGGTGACGGCCAGGAGGATGAGTTAA 192

Qy  121  AGGTGAAAATGACAAGGTTTTCCACCCCTCAAACCTTGGCTCCTTTTTCTGACAAATACAGTC  180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  193  AGGTGAAAATGACAAGGTTTCCACCCCTCAAACCTTGGCTCCCTTTTTCTGACAAATACAGTC 252

Qy  181  TGAATGAAC 189
```

RESULT 11

```
LOCUS       BD032418                261 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Sequence tag and encoded human protein.
ACCESSION   BD032418
VERSION     BD032418.1  GI:22574160
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 261)
  AUTHORS   Edwards,J.B.D.M., Duclair,E. and Jordan,J.Y.
  TITLE     Sequence tag and encoded human protein.
  JOURNAL   Patent: JP 2001269182-A 8664 02-OCT-2001;
COMMENT     GENSET
            OS   Homo sapiens (human)
            PN   JP 2001269182-A/8664
            PD   02-OCT-2001
            PF   24-FEB-2000 JP 200118773
            PR   26-FEB-1999 US  60/122487
            PI   JEAN BAPUTIST DUMAS MILNE EDWARDS,EIMERIC DUCLAIR,JEAN YVES
            PI   JORDAN
            PC   C12N15/09,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N1/21, PC
            PC   C12N5/10,
            PC   C12P21/02,C12P21/08,C12Q1/68//G06F17/30,C12N15/00,C12N5/00, PC
                 G06F15/40
            CC
            FH   Key             Location/Qualifiers
                 1..261
                 source     Location/Qualifiers
                            1..261
                            /organism="Homo sapiens"
                            /mol_type="genomic DNA"
                            /db_xref="taxon:9606"
ORIGIN

Query Match         23.0%;  Score 187;  DB 6;  Length 261;
Best Local Similarity  98.9%;  Pred. No. 9.4e-39;
Matches 187;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy    1 CACTGGAGTGGGGATGGTCCACGGGCAAACTATTAAACTGATTCTCATCAGAGAAACTGCACA 60
        ||||||| ||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   73 CACTGGANNGGGGATGGTCCATCGGGCAACTATTAAACTGATTCTCATCAGAGAAACTGCACA 132

Qy   61 TTATCTCCCCATCACTTCAAAGGTCTCGTCAGGCACAGGAGGTGACGCGCAGGAGGATGAATTTAA 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133 TTATCTCCCCATCACTTCAAAGGTCTCGTCAGGCACAGGAGGTGACGCGCAGGAGGATGAATTTAA 192

Qy  121 AGGTGAAAATGACAAGGTTTCCACCCCCTCAAACCTTGGCTCCTTTTCTGACAATACAGTC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  193 AGGTGAAAATGACAAGGTTTCCACCCCCTCAAACCTTGGCTCCTTTTCTGACAATACAGTC 252

Qy  181 TGAATGAAC 189
        |||||||||
Db  253 TGAATGAAC 261
```

```
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173288)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 173288)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Feb 7, 2002 this sequence version replaced gi:17352431.
            --------------- Genome Center
            Center: University of Washington Genome Center
            Center Code: UWGC
            Web site: http://www.genome.washington.edu
            Contact: uwgchtgs@u.washington.edu
            Drafting Center: BCM
            --------------- Project Information
            Center project name: chr-3
            Center clone name: RP11-253K11 (bc0635)
            --------------- Summary Statistics
            Sequencing vector: unknown; 50% of reads
            Sequencing vector: plasmid; 50% of reads
            Chemistry: Dye-terminator ET; 80% of reads
            Chemistry: Dye-terminator Big Dye; 20% of reads
            Assembly program: Phrap; version 0.990319
            Assembly program: Phrap; version 0.990319
            Consensus quality: 173003 bases at least Q40
            Consensus quality: 173181 bases at least Q30
            Consensus quality: 173237 bases at least Q20
            Insert size: 173288; sum-of-contigs
            Quality coverage: 10.4x in Q20 bases; sum-of-contigs
            ---------------
            Overlapping Sequences:
            5': RP11-306N11 (UWGC:bc06642) AC104443
            3': RP11-40B22 (UWGC:bc0611) AC046199
            ---------------
            Sequence Quality Assessment:
            This entry has been annotated with sequence quality
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality zero.
            Quality levels above 40 are expected to have less than
            1 error in 10,000 bp.
            Base-by-base quality values are not generally visible from the
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.
            ---------------
            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            ---------------
            Sequence Validation:
            This sequence has been validated by Multiple Complete Digest
            fingerprinting. Comparison of the experimentally derived digest
            fragments with sequence-predicted fragments is given below.
            The electronically-digested sequence consists of both insert and
            vector, in order to accurately represent the entire circular BAC.
            Small fragments below a variable cutoff (approximately 400-800 bp)
            are not resolved in the fingerprint and hence do not appear
            in the table. There are no significant remaining discrepancies
            between the experimental and predicted values. Uniquely ordered
            fragments are separated by dashed lines.

               EcoRI                 HindIII                 BglII
             -----------           -----------            -----------
            SeqDerMap FngrPrnt    SeqDerMap FngrPrnt     SeqDerMap FngrPrnt
```

RESULT 12

```
LOCUS       AC104167             173288 bp    DNA     linear   PRI 07-FEB-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-253K11, complete sequence.
ACCESSION   AC104167
VERSION     AC104167.2  GI:18583922
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173288)
```

8696	8884	3187	3244	3162	3246	4677	4625	945	883	1128	1116
6	<800	6382	6545	2057	2082	1141	1116	562	<800	829	850
1870	1827	512	5399	819	836	3303	3278	4205	4172	3107	3246
4820	4754	449	<800	1097	1116	819	836	<800	<800	3283	3246
7995	7888	15274	15715	672	<800	4124	4126	525	<800	10528	10519
142	<800	7177	2783	2823		883	888	9661	9490	2163	2249
3847	3775	2682	2895	4088	4076	2305	2394	1892	1902	269	<800
3019	2984	2876	2895	3330	3246	<800	9519	9526	1120	1140	1612	1625
6844	6834	726	<800	2082	2082	2181	2036	484	<800	588	<800
3829	3775	2563	2650	4622	4580	9684	9526	849	853	1023	1062
2135	2129	8723	8629	5562	5484	11296	11217	2509	2650	2971	2959
1430	1386	695	<800	9557	9643	234	<800	41	<800	1372	1349
115	<800	118	<800	1062	1062	6265	6174	4610	4668	4124	4076
9439	9526	3042	3069	1818	1764	3585	3775	6930	6545	1107	1116
804	836	416	<800	9870	9643	601	<800	1414	1429	228	<800
856	888	1076	988	6101	6068	2655	2672	6608	6545	5400	5484
2019	2036	4516	4487	428	<800	2683	2872	4638	4668	2274	2443
2090	2129	1634	1640	2394	2443	1598	1568	3497	3506	2413	2443
2385	2394	803	742	1155	1116	2500	2672	607	<800	1239	1236
303	<800	787	<800	1902	1922	<800	1244	1213	6733	6545	845	850
1188	1213	556	<800	5063	4970	3494	3531	3037	3069	6999	7078
3021	2984	4804	4668	699	<800						
314	<800	46	<800	2018	2082						
1459	1386	1688	1730	3256	3246						
1101	1116	2177	2209	2895	2959						
517	<800	3062	3069	609	<800						
2166	2129	6219	6140	1121	1116						
2429	2394	7221	7163	3560	3591						
2440	2538	10686	10536	7604	7678						
123	<800	271	<800	6427	6467						
109	<800	1694	1730	2443	2443						
7704	7888	3465	3506	1053	1062						
20	<800	2013	2043	1419	1418						
1404	1386	2869	2895	2619	2647						
1420	5460	5460	5437	2106	2082						

```
RESULT 13
AC112214
LOCUS       AC112214              173990 bp    DNA     linear   PRI 29-MAR-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-649F12, complete sequence.
ACCESSION   AC112214
VERSION     AC112214.2  GI:19807861
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Query Match             17.9%;  Score 145;  DB 9;  Length 173288;
Best Local Similarity   100.0%;  Pred. No. 1.5e-27;
Matches  145;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     499 CCAGAAGGAGGAGGACGTGTGAAGTGATAGACAGCACACACCGATGTTGTTAACAAGAATCGCATT 558
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76379 CCAGAAGGAGGAGGACGTGTGAAGTGATAGACAGCACACACCGATGTTGTTAACAAGAATCGCATT 76438

Qy     559 GAGGAGGCGGGTCACAAACAGTAAAGTGTTCCTGTCTTACCTGGAAAAGTGGCTGGAACAACA 618
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76439 GAGGAGGCGGGTCACAAACAGTAAAGTGTTCCTGTCTTACCTGGAAAAGTGGCTGGAACAACA 76498

Qy     619 AGAAACCGGCCTTCTTGCGTCGATG 643
           |||||||||||||||||||||||||
Db   76499 AGAAACCGGCCTTCTTGCGTCGATG 76523
```

```
REFERENCE   1  (bases 1 to 173990)
AUTHORS     Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 173990)
AUTHORS     Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (20-FEB-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 173990)
AUTHORS     Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (29-MAR-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Mar 29, 2002 this sequence version replaced gi:18767513.
            --------- Genome Center
            Center: University of Washington Genome Center
            Center Code: UWGC
            Web site: http://www.genome.washington.edu
            Contact: uwgc@gs.washington.edu
            --------- Project Information
            Center project name: chr-3
            Center clone name: RP11-649F12 (bc0666)
            --------- Summary Statistics
            Sequencing vector: Plasmid; 100% of reads
            Chemistry: Dye-terminator ET; 54% of reads
            Chemistry: Dye-terminator Big Dye; 46% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 173825 bases at least Q40
            Consensus quality: 173976 bases at least Q30
            Consensus quality: 173990 bases at least Q20
            Insert size: 173990; sum-of-contigs
            Quality coverage: 7.6x in Q20 bases; sum-of-contigs
            --------- Overlapping Sequences:
            5': RP11-306N11 (UWGC:bc0642) AC104443; 91913-bp overlap
            3': RP11-253K11 (UWGC:bc0635) AC104167; 126067-bp overlap
            --------- Sequence Quality Assessment:
            This entry has been annotated with sequence quality
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality zero.
            Quality levels above 40 are expected to have less than
            1 error in 10,000 bp.
            Base-by-base quality values are not generally visible from the
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.
            --------- Sequence Validation:
            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            --------- Sequence Validation:
            This sequence has been validated by Multiple Complete Digest
            fingerprinting. Comparison of the experimentally derived digest
            fragments with sequence-predicted fragments is given below.
            The electronically-digested sequence consists of both insert and
            vector, in order to accurately represent the entire circular BAC.
            Small fragments below a variable cutoff (approximately 400-800 bp)
            are not resolved in the fingerprint and hence do not appear
            in the table. There are no significant remaining discrepancies
            between the experimental and predicted values. Uniquely ordered
            fragments are separated by dashed lines.
```

EcoRI		BglII		HindIII	
SeqDerMap	FngrPrint	SeqDerMap	FngrPrint	SeqDerMap	FngrPrint
8696	8565	6884	7019	3687	3659
----	<800	2067	2033	6382	6611
7665	7607	8943	9146	512	<800
6	<800				
4890	4847	204	<800	449	<800
2632	2686	966	979	11952	12067
8473	8565	669	<800	4600	4499
176	10	<800	<800	1352	1316
7	6231	6122	342	<800	
2108	2145	3467	3391	8281	8266
791	813	1208	1186	48	<800
3999	4106	862	849	5601	5476
2338	2442	1928	2033	1641	1690
45	<800	11992	11986	1773	1690
2558	2442	7232	7019	3086	3129
12241	12505	1436	1398	83	<800
3573	3584	301	<800	3422	3478
2651	2686	436	<800	950	956
3795	3976	2042	2033	1599	1556
656	<800	1201	1186	935	956
3738	3717	3199	3197	388	<800
1378	1368	954	979	2059	2189
2901	2904	1778	1745	3684	3659
3487	3497	1454	1398	2170	2189
432	<800	197	<800	3037	3027
1244	1216	1677	1612	2765	2770
2498	2553	5157	5103	6731	6611
2024	2002	6	<800	3647	3659
2655	2814	6985	7019	983	956
2692	2686	1780	1745	3496	3478
1600	1557	4390	4362	600	<800
6264	6067	1239	1186	8344	8266
234	<800	2415	2409	4609	4499
2162	2145	843	849	6619	6611
11293	11369	2283	2409	41	<800
9673	9507	5400	5307	2509	2570
2181	2145	228	849	849	807

9522	---	9507			1107	1112	<800
2248	---	2261	4124	4049	1120	1122
883	---	898	1371	1398	1889	1920
4124	---	4106	2971	3081	9661	9610
819	---	813	1023	979	525	<800
3304	---	3334	588	315	<800
1141	---	1109	1612	1612	4194	4146
4677	---	4566	269	<800	562	<800
7691	---	7607	3107	3197	1956	2031
109	---	<800	829	849	3465	3478
1404	---	1368	10528	10505	5463	5476
20	---	<800	3272	3197	2869	2869
123	---	<800	1131	1112	1694	1690
2440	---	2442	2106	2033	271	<800
2429	---	2553	2619	2622	10687	10712
2168	---	2145	1419	1398	7222	7247
517	---	<800	996	979	6205	6158
1101	---	1109	2443	2409	3064	3027
1459	---	1368	6429	6428	2177	2189
314	---	<800	7603	7648	1688	1690
4210	---	4106	3560	3557	46	<800

Query Match   17.9%;   Score 145;   DB 9;   Length 173990;
Best Local Similarity 100.0%;   Pred. No. 1.5e-27;
Matches 145;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
Qy     499 CCAGAAGGAGGAGGGACGTGTGAAGTAGTAGCAGCACAGCACCGATGTGTTGTAACAAGAATCGGCATT 558
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124282 CCAGAAGGAGGAGGGACGTGTGAAGTAGTAGCAGCACAGCACCGATGTGTTGTAACAAGAATCGGCATT 124341

Qy     559 GAGGAGCGGGTCACAAACAGTAAAAGTGTTCCTGTCTACCTGGAAAAAGTGGGACAACAAGAAACCG 618
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124342 GAGGAGCGGGTCACAAACAGTAAAAGTGTTCCTGTCTACCTGGAAAAAGTGGGACAACAAGAAACCG 124401

Qy     619 AGAAAACCGGCCGGTCACAAAACAGTTCCTTCTTCGGTCGGATG 643
           ||||||||||||||        ||||||||||||||
Db  124402 AGAAAACCGGCCGCTTCTTCTTCGGTCGGATG 124426
```

RESULT 14
LOCUS   AY325123   901 bp   mRNA   linear   ROD 23-JUL-2003
DEFINITION Mus musculus TAFA4 mRNA, complete cds.
ACCESSION   AY325123
VERSION   AY325123.1 GI:32967244
KEYWORDS   .
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 901)
  AUTHORS   Tang,Y.T., Emtage,P., Funk,W., Hu,T., Arterburn,M., Park,E. and
            Rupp,F.
  TITLE   TAFA: A Novel Secreted Family with Homology to CC-chemokines
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 901)
  AUTHORS   Tang,Y.T., Emtage,P., Funk,W., Hu,T., Arterburn,M., Park,E. and
            Rupp,F.
  TITLE   Direct Submission
  JOURNAL   Submitted (17-JUN-2003) Biology Research, Nuvelo, 675 Almanor
            Avenue, Sunnyvale, CA 94085, USA
FEATURES   Location/Qualifiers
     source   1..901
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
     CDS   206..613
                     /note="secreted protein similar to CC-chemokines"
                     /codon_start=1
                     /product="TAFA4"
                     /protein_id="AAP92415.1"
                     /db_xref="GI:32967245"
                     /translation="MRVCAKVLLSRWLVLTYVLMVYCCKLMSASSQHLRGHAGHHLIK
                     PGTCEVVAVHRCCNKNRIEERSQTVKCSCFPGQVAGTTRAQPSCVERAIVIEKWWCHM
                     NPCLEGEDCKVLPDSSGWSCSSGNKVKTTKVTR"
ORIGIN

RESULT 15
LOCUS   AK057890   1897 bp   mRNA   linear   PRI 12-SEP-2003
DEFINITION Homo sapiens cDNA FLJ25161 fis, clone CBR08081.
ACCESSION   AK057890
VERSION   AK057890.1 GI:16553863
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,
            Fukuzumi,Y., Fujimori,Y., Komiyama,M., Suzuki,Y., Hata,H.,
            Nakagawa,K., Mizuno,S., Morinaga,M., Kawamura,M., Sugiyama,T.,
            Irie,R., Otsuki,T., Sato,H., Nishikawa,T., Sugiyama,A.,
            Kawakami,B., Nagai,K., Isogai,T. and Sugano,S.
  TITLE   NEDO human cDNA sequencing project
  JOURNAL   Unpublished

Query Match   16.9%;   Score 137.4;   DB 10;   Length 901;
Best Local Similarity   67.0%;   Pred. No. 1.5e-25;
Matches 195;   Conservative 0;   Mismatches 96;   Indels 0;   Gaps 0;

```
Qy  507 AGGGACGTGTGAAGTGATAGCAGCACACCGATGTGTTGTAACAAGAATCGGCATTGAGGAGCG 566
        ||||| || | |||| || | | || || | ||| |||| ||||||||| | || ||||||||
Db  340 AGGGTACCTCCGAGTGGTGGTCGGCCGTGCACCGGCTGCTGCAATAAGAATCCGGATTAGGAGGAGCG 399

Qy  567 GTCACAAACAGTAAAAGTGTTCCTGTCTACCTGGAAAAAGTGGGACAACAAGAAACCG 626
        ||| | || |||| |||||||||| | |   | || |  | | || | | |||  |||
Db  400 CTCTCAGACACGGTCAAAATGTTCCTTCCTTCCCAGGCCGACGGTGGCGTGGGCCACA 459

Qy  627 GCCTTCTTCGGTCGTCGATGCCTCCATTGTTGTGTAGTAGTTGGGAAAATGGTGGTGTGAGATGGAGCCTTG 686
        ||||||||  | ||  |  | |||||||||| | |   |||  |||  | ||||||| | || ||||||||||
Db  460 GCCTTCTTGTTGTTGAAGCTCCCATTGTGATTGAGGAGAGTGGTGGTGCCCACATGAATCCTTG 519

Qy  687 CCTAGAAGGAGGAAGAAATGTAAGACACTCCCTGACGAATTCTGGGATGGAATGTGCGCGCAACAGG 746
        | | ||||||||| |  | ||||||  |  ||||||| | |||| |||  || |||  | | ||| ||
Db  520 CTTGGAAGGAGGAGGACTGTAAAAGTGGCTTCCAGGACTCTTCGGGGTTGGTCTGTACCAGTGG 579

Qy  747 CAACAAAAATTAAGACCACGAGAATTCACCCAAGAACCTAACAGAGAAGCATTT 797
        ||| ||||| | |||  ||| |    |||| | |  ||||  || ||| |||||
Db  580 CAATAAAAGTCAAAAACCACAAGGTGACCCGGTAGCGAGACACACAAAAGTGTTT 630
```

```
REFERENCE   2 (bases 1 to 1897)
AUTHORS     Sugano,S. and Suzuki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (24-OCT-2001) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Laboratory of Genome Structure, Human Genome
            Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
            (E-mail:flcdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA full insert
            construction and 5'-end one pass sequencing: Institute of Medical
            Science, University of Tokyo, Laboratory of Genome Structure, Human
            Genome Center; 3'-end one pass sequencing: RAB; clone selection for
            full insert sequencing: RAB and Helix Research Institute.
FEATURES            Location/Qualifiers
     source         1..1897
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="CBR08081"
                    /tissue_type="brain"
                    /clone_lib="CBR"
                    /note="cloning vector: pME18SFL3"
     CDS            110..532
                    /note="unnamed protein product"
                    /codon_start=1
                    /protein_id="BAB71606.1"
                    /db_xref="GI:16553864"
                    /translation="MRSPRMRVCAKSVLLSHWLFLAYLMVCCKLMSASSQHLRGHAG
                    HHQIKQGTEVAVAHRCCNKDNRIEERSQTVKCSCFPGQVAGTTRAQPSCVEASIVIQK
                    WWCHNNPCLEGEDCKVLPDYSGWSCSSGNKVKTTKVTR"
ORIGIN
Query Match     16.8%;  Score 136.8; DB 9;  Length 1897;
Best Local Similarity   70.4%;  Pred. No. 2.2e-25;
Matches  183;  Conservative   0;  Mismatches  77;  Indels   0;  Gaps   0;

Qy    507 AGGGACAGCGTGTGAAGTGATAGCAGCACACCGATGCTTGTTAACAAGAAATCGCATTGAGGAGCG 566
          ||||||||  |||  ||  |||||||||  |  ||| || |||| |||| ||  |||| ||||||||
Db    259 AGGGACCTGTGAGGTGGTCGCGCCGTGCACAGGTGCTCGCAATAAGAACCGCATAGAGACAGAGCG 318

Qy    567 GTCACAAACAGTAAAAGTGTTCCTTGTCCTTACCTGGGAAAAGTGGCTGGACACAACAAGAAAACCG 626
          ||||||| ||  ||| | |||| |||| |||  |||| |||||||||| ||||  || |||| |||
Db    319 GTCACAAAACGGTTCAAGTTGCTCTTTGCTTCCCCGGGACAGGTGGCGGGGCACAACTCGGGCTCA 378

Qy    627 GCCTTCTTGCGTCGATGCCTTCCATAGTGATTGGGAAAATGGTGGTTGTGAGATGGGAGGCCTTG 686
          |||||||| | ||||  |||||  ||||| |||| ||| ||||| ||| ||||| |||| |||||
Db    379 ACCTTCTTGTGTTGAACCTTCCATTGTGATTCAGAAAATGGGTGGTGTCACATGAAATCCGTG 438

Qy    687 CCTAGAAGGAGGAAGAATTAAGACACTCCCTGACAATTCTGGATTGGATGTGGGCAACAGG 746
          |||||  || || |  ||||  || | |  |||||||| || ||| || |||||| ||| ||
Db    439 TTTGGAAGGAGGAAGAATTGTAAAGTGCTGGCCAGATTACTCAGGTTTGGTCCCTGTAGCAGTGG 498

Qy    747 CAACAAAATTAAGACCACGA 766
          ||| ||  ||||  |||| ||
Db    499 CAATAAAGTCAAAAACTACGA 518
```

Search completed: April 9, 2004, 14:55:30
Job time : 3299 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: April 9, 2004, 12:56:28 ; Search time 94 Seconds
(without alignments)
4793.831 Million cell updates/sec

Title: US-09-763-335-4
Perfect score: 812
Sequence: 1 cactgggagtggggatggtcc.........catttgtcgggtagtaaaggaa 812
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Database :
Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match Length	DB	Match Length	DB ID	Description
1	131.2	16.2	4	716	US-09-620-312D-225	Sequence 225, App
2	88.8	10.9	3	333	US-08-905-223-218	Sequence 218, Appl
3	46.4	5.7	1	7218	US-08-232-463-14	Sequence 14, Appl
4	44.4	5.5	4	832	US-09-621-976-2813	Sequence 2813, Ap
5	39	4.8	3	832	US-09-621-976-2813	Sequence 2813, Ap
6	38.8	4.8	4	2060	US-09-370-807-1	Sequence 1, Appl
7	38.6	4.8	3	2060	US-09-921-259-1	Sequence 1, Appl
8	37	4.6	1	2914	US-08-454-097-11	Sequence 11, Appl
9	37	4.6	3	2914	US-09-921-259-11	Sequence 11, Appl
10	37	4.6	4	6866	US-10-204-708-19	Sequence 19, Appl
11	36.2	4.5	3	2143	US-09-673-395A-82	Sequence 82, Appl
12	36.2	4.5	4	3174	US-08-545-528D-1	Sequence 1, Appl
13	36.2	4.5	4	580073	US-08-489-847-86	Sequence 86, Appl
14	35.8	4.4	4	472	US-09-439-313-351	Sequence 351, Appli
15	35.8	4.4	4	472	US-09-352-616A-351	Sequence 351, App
16	35.8	4.4	4	472	US-09-636-215-351	Sequence 351, App
17	35.4	4.4	4	472	US-09-685-166A-351	Sequence 351, App
18	35.4	4.4	3	4147	US-09-688-078-6	Sequence 6, Appli
19	35	4.3	4	1000	US-09-641-638-357	Sequence 357, App
20	35	4.3	4	5177	US-09-568-407-2	Sequence 2, Appl
21	34.6	4.3	4	406	US-09-702-705-1150	Sequence 1150, Ap
22	34.6	4.3	4	406	US-09-736-457-1150	Sequence 1150, Ap
23	34.6	4.3	4	406	US-09-614-124B-1150	Sequence 1150, Ap
24	34.6	4.3	4	406	US-09-671-325-1150	Sequence 1150, Ap
25	34.6	4.3	5	5445	US-09-689-055B-2	Sequence 2, Appli
26	34.6	4.3	1	640681	US-09-790-988-1	Sequence 1, Appli
27	34.4	4.2	4	696	US-09-134-001C-1287	Sequence 1287, Ap
28	34.2	4.2	4	571	US-08-858-207A-262	Sequence 262, App
29	34.2	4.2	4	3867	US-09-366-715-5	Sequence 5, Appli
30	34.2	4.2	4	5365	US-08-961-527-77	Sequence 77, Appl
31	33.8	4.2	4	1001	US-09-641-638-356	Sequence 356, App
32	33.8	4.2	4	3824	US-09-326-529-3	Sequence 3, Appli
33	33.8	4.2	4	13086	US-08-956-171E-16	Sequence 16, Appli
34	33.8	4.2		162450	US-09-345-882-1	Sequence 1, Appl
35	33.4	4.1	4	301	US-09-439-313-263	Sequence 263, App
36	33.4	4.1	4	301	US-09-439-313-302	Sequence 302, App
37	33.4	4.1	4	301	US-09-352-616A-263	Sequence 263, App
38	33.4	4.1	4	301	US-09-352-616A-302	Sequence 302, App
39	33.4	4.1	4	301	US-09-232-149A-263	Sequence 263, App
40	33.4	4.1	4	301	US-09-232-149A-302	Sequence 302, App
41	33.4	4.1	4	301	US-09-159-812-263	Sequence 263, App
42	33.4	4.1	4	301	US-09-159-812-302	Sequence 302, App
43	33.4	4.1	4	301	US-09-636-215-263	Sequence 263, App
44	33.4	4.1	4	301	US-09-636-215-302	Sequence 302, App
45	33.4	4.1	4	301	US-09-685-166A-263	Sequence 263, App

ALIGNMENTS

RESULT 1
US-09-620-312D-225
Sequence 225, Application US/09620312D
Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Polypeptides
; TITLE OF INVENTION: No. 6569662e1 Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 225
; LENGTH: 716
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(596)
US-09-620-312D-225

Query Match 16.2%;
Best Local Similarity 63.1%;   Score 131.2; DB 4; Length 716;
Matches 202; Conservative 0; Mismatches 118; Indels 0; Gaps 0;   Pred. No. 9.8e-32;

```
Qy  479 AGCAGCATCACCTGCACAGACCACAGAAGGAGGCAGGTGTGAAGTGATAGAGCACACCGAT 538
            ||| ||||| ||| ||||  || ||| || ||  | ||| |||| || ||  ||||
Db  295 ACCATCATAAAGTCTCCACCATGTTAAAACGGGAACTTGTGAGGTGGTGGCACTCCACAGAT 354
```

Qy 539 GTTGTAACAAGAAATCGCAATTGAGGAGGCGGTCACAAACAGTAAAAGTGTTCCTGTCTACCTG 598
Db 355 GCTGTAATAAGAACAAGATAGAGAAGGACGGTCACAACAGTCAAGTGCTTCCCTGCTTCCCTG 414

Qy 599 GAAAAGTGGCGTGGAACAACAGAAAACGGGCCTTTCTGCGTGGATGCCTCCATAGTGGATTG 658
Db 415 GGCAGGTGGGCAGGCACCACGCGAGCTGCTCCATCATGTGGATGCTTCAATAGTGGAAC 474

Qy 659 GGAAAATGGTGGTGTGAGGATGGGAGCCTTGCCTTAGAAGGAGGAGAAGAATGTAAGACACTCCCTG 718
Db 475 AGAAAATGGTGGTGTGCCAATATGCAGCCCATGTCTAGAGGGGGAGAAGAATGTAAAGTTCTTCCGG 534

Qy 719 ACAATTCTGGATGGATGGTGCGCAACAGCGACACAAAATTAAGACCACGAGAATTCACCAA 778
Db 535 ATCGGAAAAGGATGGGAGCTGTTCCTCTCGGGAATAAAAGTCAAAAACAACTAGGGTAACCCATT 594

Qy 779 GAACCTAACACAGAAGCATTTG 798
Db 595 AACCCAGAGAGAAATCAAGTG 614

RESULT 2
US-08-905-223-218
Sequence 218, Application US/08905223
Patent No. 6222029
GENERAL INFORMATION:
APPLICANT: Edwards, Jean-Baptiste D.
APPLICANT: Duelert, Aymeric
APPLICANT: Lacroix, Bruno
TITLE OF INVENTION: 5' ESTs FOR SECRETED PROTEINS
NUMBER OF SEQUENCES: 503
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 501 West Broadway
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-3505
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Win95
SOFTWARE: Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/905,223
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 218:
SEQUENCE CHARACTERISTICS:
LENGTH: 333 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
TISSUE TYPE: Brain
FEATURE:
NAME/KEY: other
LOCATION: 32..331
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 1..300
OTHER INFORMATION: id R13004
OTHER INFORMATION: est
FEATURE:

FEATURE:
NAME/KEY: other
LOCATION: 114..274
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 54..214
OTHER INFORMATION: id T80337
OTHER INFORMATION: est
FEATURE:
NAME/KEY: other
LOCATION: 272..331
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 213..272
OTHER INFORMATION: id T80337
OTHER INFORMATION: est
FEATURE:
NAME/KEY: other
LOCATION: 66..106
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 6..46
OTHER INFORMATION: id T80337
OTHER INFORMATION: est
FEATURE:
NAME/KEY: other
LOCATION: 101..278
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 94
OTHER INFORMATION: region 70..247
OTHER INFORMATION: id T08840
OTHER INFORMATION: est
FEATURE:
NAME/KEY: other
LOCATION: 101..249
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 95
OTHER INFORMATION: region 72..220
OTHER INFORMATION: id HSCOCF041
OTHER INFORMATION: est
FEATURE:
NAME/KEY: other
LOCATION: 33..113
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 98
OTHER INFORMATION: region 1..81
OTHER INFORMATION: id T08840
OTHER INFORMATION: est
FEATURE:
NAME/KEY: other
LOCATION: 31..112
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 1..82
OTHER INFORMATION: id HSCOCF041
OTHER INFORMATION: est
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 247..321
IDENTIFICATION METHOD: Von Heijne matrix
OTHER INFORMATION: score 5.8
OTHER INFORMATION: seq LWISACAMLLCHG/SL
US-08-905-223-218

Query Match 10.9%; Score 88.8; DB 3; Length 333;
Best Local Similarity 97.8%; Pred. No. 2e-18;
Matches 90; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 378 TCAGAGAATGGCAATAGGTGTTCCTCGGATGGTCCTGGGTTCCTGGTTCCTGTATTTGTGGATAAGTGCTTTG 437
Db 240 TTATAGAATGGCAATGGTGTCTCTGCGATGGTGTCCTCCTGGGTTCCTGTATTTGTGGATAAGTGCTTG 299

Qy 438 TGCAATGCTTACTCTGCCATGGATCCCTTCAGC 469

Db   300 TGCAATGCTACTCTGCCATGGATCCCTTCAGC 331

RESULT 3
US-08-232-463-14/c
Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; PRIOR APPLICATION DATA:
; CLASSIFICATION:
; APPLICATION NUMBER: 435
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pIzgpt-F1s
; US-08-232-463-14

Query Match   5.7%; Score 46.4; DB 1; Length 7218;
Best Local Similarity 2.4%; Pred. No. 0.0039;
Matches   8; Conservative 195; Mismatches 131; Indels 0; Gaps 0;

Qy   479 AGCAGCATCACCTGCACAGACCAGAAGGAGGAGGACGTGTGAAGTGATAGCAGCACACCGGAT 538
Db 1449 AGAAGAAATTCGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1390
Qy   539 GTTGTAACAAGAATCGCCATTGAGGAGCGGGTCACAAACAGTAAAGTGTTCCTGTCTTACCTG 598
Db 1389 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1330
Qy   599 GAAAAGTGGCTGGACAACAAGAAACCGGCCTTCTTGCGTCGATGCCTCCATAGTGATG 658
Db 1329 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1270
Qy   659 GGAAATGGTGGTGTGAGATGGAGCCTTGCCTAGGAGGAGAAGAATGTAAGACACTCCCTG 718
Db 1269 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1210
Qy   719 ACAATTCTGGATGGGATGGAGTGCGGCCACAGGCAACAAAAATTAAGACCACGAGAATTCACCCAA 778

Db   1209 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1150
Qy   779 GAACCTAACCAGAAGCATTGTGGTAGTAAAGGAA 812
Db   1149 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1116

RESULT 4
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match   5.5%; Score 44.4; DB 4; Length 832;
Best Local Similarity 11.4%; Pred. No. 0.00048;
Matches 27; Conservative 119; Mismatches 90; Indels 0; Gaps 0;

Qy   57 CACATTATCTCCCCATCACTTCAAAGGTCTGTCGCAGGCAGAGGTGACGCCAGGAGATGAT 116
Db   237 CATWCYWYWYXYWKRMWSKTCWSGSRGGYWTSYTSTRSYSMYWASWMYTMCWWWGRWWSTY 178
Qy   117 TTAAAGGTGAAAAATGACAAGGTTTCCACCCCTCAAAACTTTGGCTCTTTCTGACAATAC 176
Db   177 WYMANGKKKWWRYATTWRRAMMWWWAAWTWMWYMWWAWCMWSSRGAAMYRRTMMWGYRYWW 118
Qy   177 AGTCTGAATGAACCCGATGTCTTTTTTTTTTTACTGTGGAAATAGGAATCGGAAGAGAGTAAC 236
Db   117 RKKSYRRTRCAWAYAWKIKRSYTWCWRWKWKRCMMMMMMMAMAYGKTMMWRACWKTRYRRW 58
Qy   237 ATTTTTTTTTTAATCCTGATAAAGAAAGATTGTTGGGAAGCTCTTTGAAAAAAATT 292
Db   57 WAWAMWRXWWTMMMYYWTWRAMXRWWMWRKRWSWSWWMAWGWTRWAARMWWRWY 2

RESULT 5
US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399

US-09-621-976-2813

Query Match   4.8%; Score 39; DB 4; Length 832;
Best Local Similarity   10.4%; Pred. No. 0.025;
Matches 33; Conservative 156; Mismatches 126; Indels   2; Gaps   1;

Qy   196 TCTTTTTTTTTTACTTGTGGGAAATAGGATCGGAAGGAGGAGTAACATTTTTTTTTTAATCCTGA 255
           :::::|:::::|:::::|:::|::::::::::::::::::::::::::::::::|::|:::::
Db   7 KYTTWTAKCWTKWKWKNSWSYWMWKWYMYYMKTYWRWRRKKKKAWWKYWKTWTWWWYWRYAMVG   66

Qy   256 TAAAGAAGAATTGTTGGGAAGCTTCTTTGAAAAAAAAAAATTTTAAATTGTGGCACAGATGGATT 315
           :::::::::::::::::|:::::::::::|:|:::::::::::::::::::::::::::::|::::
Db   67 TYKKKAMCRTKTKKKKGIYMWMWTWGWRRSYMAWWTRWTGYAYYRSMMYWYWYRCWKK   126

Qy   316 TTTAAAAGTGTTTAGATCTTTCCAATGAAACACTAATAGAGTACTCTGCTCTTTGGCTGGATT 375
           :::::::::::::::::::::::|::.:::::::::::::::::::|::::::: :: :::
Db   127 KAYYRKTTCYSSKGWTWWKRWKKAWTTWWWKKIYTWAAATRYWWMMCWTKRWRASWWYCWW   186

Qy   376 TTTCACAG--AATGGCAAATGGTCTCTGCGCGATTGTCCTGGGTYCCTGTATTTGTGGATAAGTG 433
           :::::   ::|:::::::::::::::::::::::|:|:::|::|:::
Db   187 WGKARKWSTWRKSRSYASARSAKRCCYSCSWGAMSWKYMWRMWRWRGWWATGAGMKAWRAS   246

Qy   434 CTTGTGCAAATGCTACTCTGCCATGGAATCCCCTTCAGCACACTTTCCAGCAGCGCATCACCTGC 493
           :::::::|:::|:::::|:::::::: :|:::|:: |:|:|:::
Db   247 CMMRRXYAGKSKTSYXSMWMCWTRSWKYCYTKARWTGYYCYRKGGWWGKRGRWYASKKYM   306

Qy   494 ACAGACCAGAGAGGAGGG 510
           :::::::::::::::::
Db   307 WXRWWWCWARMYRYSTG 323

RESULT 6
US-09-370-807-1/c
Sequence 1, Application US/09370807
Patent No. 6297034
GENERAL INFORMATION:
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Falco, S. Carl
APPLICANT: Rafalski, J. Antoni
APPLICANT: Sakai, Hajime
TITLE OF INVENTION: N-End Rule Pathway Enzymes
FILE REFERENCE: BB-1199
CURRENT APPLICATION NUMBER: US/09/370,807
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: 60/096,225
EARLIER FILING DATE: August 12, 1998
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Microsoft Office 97
SEQ ID NO 1
LENGTH: 2060
TYPE: DNA
ORGANISM: Glycine max
US-09-370-807-1

Query Match   4.8%; Score 38.6; DB 3; Length 2060;
Best Local Similarity   54.6%; Pred. No. 0.057;
Matches 77; Conservative   0; Mismatches 64; Indels   0; Gaps   0;

Qy   238 TTTTTTTTTTTTTAATCCTGATAAAAGAAGAATTGTTGGGAAGCTTCTTTGAAAAAAAAAATTTAAA 297
           |||||||||| ||| |||||||||||||||||||||||||||
Db   2058 TTTTTTTTTTTTTTTTTGTAAAAAAAGAATATGTAAAAAGTGTCATTAACCAAAAAACTTATGG 1999

Qy   298 TTGTGGCACAGATGGATTTAGATCTTTCCAATGAAACACTAATAGAGTAC 357
           |||||||||| |||||
Db   1998 GTTCACAACAATTGCTATTGCCAACATTGTCAGACCATTCAAAGGAAGAATACATGAATCT 1939

Qy   358 TCTGCTCTTTGGCTGGATTTT 378
           ||||||||||||
Db   1938 TCTTTGCAGAGTTTTTTTTTT 1918

RESULT 7
US-09-921-259-1/c
Sequence 1, Application US/09921259
Patent No. 6465234
GENERAL INFORMATION:
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Falco, S. Carl
APPLICANT: Rafalski, J. Antoni
APPLICANT: Sakai, Hajime
TITLE OF INVENTION: N-End Rule Pathway Enzymes
FILE REFERENCE: BB-1199
CURRENT APPLICATION NUMBER: US/09/921,259
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 60/096,225
PRIOR FILING DATE: August 12, 1998
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Microsoft Office 97
SEQ ID NO 1
LENGTH: 2060
TYPE: DNA
ORGANISM: Glycine max
US-09-921-259-1

Query Match   4.8%; Score 38.6; DB 4; Length 2060;
Best Local Similarity   54.6%; Pred. No. 0.057; Mismatches 64; Indels   0; Gaps   0;
Matches 77; Conservative   0;

Qy   238 TTTTTTTTTTTTTAATCCTGATAAAAGAAGAATTGTTGGGAAGCTTCTTTGAAAAAAAAAATTTAAA 297
           |||||||||| ||| |||||||||||||||||||||||||||
Db   2058 TTTTTTTTTTTTTTTTTGTAAAAAAAGAATATGTAAAAAGTGTCATTAACCAAAAAACTTATGG 1999

Qy   298 TTGTGGCACAGATGGATTTAGATCTTTCCAATGAAACACTAATAGAGTAC 357
           |||||||||| |||||
Db   1998 GTTCACAACAATTGCTATTGCCAACATTGTCAGACCATTCAAAGGAAGAATACATGAATCT 1939

Qy   358 TCTGCTCTTTGGCTGGATTTT 378
           ||||||||||||
Db   1938 TCTTTGCAGAGTTTTTTTTTT 1918

RESULT 8
US-08-454-097-11/c
Sequence 11, Application US/08454097
Patent No. 5686412
GENERAL INFORMATION:
APPLICANT: Hoekstra, Merl F.
TITLE OF INVENTION: Protein Kinases
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/454,097
FILING DATE: 30-MAY-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/185,359
FILING DATE: 21-JAN-1994
APPLICATION NUMBER: US 08/008,001
FILING DATE: 21-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/728,783
FILING DATE: 03-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: No. 568641and, Greta E.
REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 27866/31853
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 312-474-6300
;   TELEFAX: 312-474-0448
;   TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 11:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2914 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 265.1275
US-08-454-097-11

Query Match 4.6%; Score 37; DB 1; Length 2914;
Best Local Similarity 50.9%; Pred. No. 0.22;
Matches 88; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Qy 181 TGAATGAACCCGATGTCTTTTTTTTTTTTTTTACTGTGGAAATAGGATCGGAAGAGAGTAACATTT 240
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2193 TGAGTCAGCAAGGCATTTTTGTTGTTTAAAAAAAATCATTTCCTTACAGAGAAACAGTT 2134

Qy 241 TTTTTTAATCCTGATAAAGAAGAATTGTTGGGAAGCTCTTTGAAAAAAAAATTTTAAAATGTTGGCA 300
    ||||||||||||||||||||||||||||||||||||
Db 2133 TTTAGTTTTTTAATGAACTTGTAAACAAAAAAAGCTCCATTTCAAAATAAAAACAAAATCC 2074

Qy 301 TGGCACAGATGGGATTTTAAAAAGTGTTAGATCTTTCCAATGAAACTAATAGA 353
    |||||||||||||||
Db 2073 CAGATCATATAGATGTTTACAGTGATTACATTTATCTAAGCAACATACATACA 2021

RESULT 9
US-08-185-359-11/c
; Sequence 11, Application US/08185359
; Patent No. 6060296
; GENERAL INFORMATION:
; APPLICANT: Hoekstra, Merl F.
; TITLE OF INVENTION: Protein Kinases
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;   STREET: 233 South Wacker Drive, 6300 Sears Tower
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: USA
;   ZIP: 60606-6402
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/185,359
;   FILING DATE:
; PRIOR APPLICATION DATA:
;   CLASSIFICATION: 435
;   APPLICATION NUMBER: US 08/008,001
;   FILING DATE: 21-JAN-1993
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/728,783
;   FILING DATE: 03-JUL-1991
; ATTORNEY/AGENT INFORMATION:
;   NAME: No. 606029and, Greta E.
;   REGISTRATION NUMBER: 35,302
;   REFERENCE/DOCKET NUMBER: 27866/31853
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 312-474-6300
;   TELEFAX: 312-474-0448
;   TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:
;   LENGTH: 2914 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 265.1275
US-08-185-359-11

Query Match 4.6%; Score 37; DB 3; Length 2914;
Best Local Similarity 50.9%; Pred. No. 0.22;
Matches 88; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Qy 181 TGAATGAACCCGATGTCTTTTTTTTTTTTTTTACTGTGGAAATAGGATCGGAAGAGAGTAACATTT 240
    ||||||||||||||||||||||||||||||||||||||||||||||||
Db 2193 TGAGTCAGCAAGGCATTTTTGTTGTTTAAAAAAAATCATTTCCTTACAGAGAAACAGTT 2134

Qy 241 TTTTTTAATCCTGATAAAGAAGAATTGTTGGGAAGCTCTTTGAAAAAAAAATTTTAAAATG 300
    ||||||||||||||||||||||||||||||||||||
Db 2133 TTTAGTTTTTTAATGAACTTGTAAACAAAAAAAGCTCCATTTCAAAATAAAAACAAAATCC 2074

Qy 301 TGGCACAGATGGGATTTTAAAAAGTGTTAGATCTTTCCAATGAAACTAATAGA 353
    |||||||||||||||
Db 2073 CAGATCATATAGATGTTTACAGTGATTACATTTATCTAAGCAACATACATACA 2021

RESULT 10
US-10-204-708-19
; Sequence 19, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10048826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 19
;   LENGTH: 6866
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-19

Query Match 4.6%; Score 37; DB 4; Length 6866;
Best Local Similarity 51.5%; Pred. No. 0.37;
Matches 85; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 186 GAACCCGATGTCTTTTTTTTTTTTTTTACTGTGGAAATAGGATCGGAAGAGAGTAACATTTTTTTTT 245
    |||||||||||||||||||||||||||||||||||||||||||||
Db 3859 GAAATAGATTAAATGTTTTTTAATAAGTAATTAGGTAAGTATATATTTATAATAATGGAAATT 3918

Qy 246 TTAATCCTGATAAAGAAGAATTGTTGGGAAGCTCTTTGAAAAAAAAATTTTAAAATGTTGGCA 305
    |||||||||||||||||||||||||||||
Db 3919 TTATTAGTTATATAAAAGAGATTTTGTAAAAAGAATATATGAAATAATTAATGTTTAGAGGTTTGT 3978

Qy 306 CAGATGGGGATTTTAAAAAGTGTTAGATCTTTCCAATGAAACACTAAT 350

```
Db    3979 TAAATGTTAATGATAAAGTTGGGGATTAAATTAGGGATTAAT 4023
              ||||  |||| |||||  ||||          |||||||| ||||
```

RESULT 11
US-09-673-395A-82/c
Sequence 82, Application US/09673395A
Patent No. 6620923
GENERAL INFORMATION:
; APPLICANT: SPECHT, THOMAS
; APPLICANT: HINZMANN, BERND
; APPLICANT: SCHMITT, ARMIN
; APPLICANT: PILARSKY, CHRISTIAN
; APPLICANT: DAHL, EDGAR
; APPLICANT: ROSENTHAL, ANDRE
; TITLE OF INVENTION: HUMAN NUCLEIC ACID SEQUENCES FROM UTERUS TUMOR TISSUE
; FILE REFERENCE: A1BRE-12
; CURRENT APPLICATION NUMBER: US/09/673,395A
; CURRENT FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 637
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 82
; LENGTH: 2143
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-673-395A-82

Query Match   4.5%; Score 36.2; DB 4; Length 2143;
Best Local Similarity   52.3%; Pred. No. 0.34;
Matches 80; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```
Qy   224 GGAAGAGAGTAACACATTTTTTTTTTTTAATCCTGATAAAGAAGAATTGTTGGGAAGCTCTTTGA 283
Db   736 GGAATACAATGGTTGTTGTTTTATCATAGTGTACACATTTAGCTTGTGTAAATGACTCACA 677

Qy   284 AAAAAAATTTAAATTGTGGCACAGATGGGGATTTTAAAAAAGTGTTAGATCTTTCCAATGAA 343
Db   676 AAACTGATTTTAAAATCAAGTTAATGTGAAATTTTGAAAAATTACTACTTAATCCTAATTCA 617

Qy   344 CACTAATAGAGAGTACTCTGCTCCTTGGCTGGGATTT 376
Db   616 CAATAACAATGGCCATTAAGGTTTGACTTGAGTT 584
```

RESULT 12
US-09-489-847-86
Sequence 86, Application US/09489847
Patent No. 6476195
GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: PZ031P1
; CURRENT APPLICATION NUMBER: US/09/489,847
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: 60/095,486
; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 60/095,657
; EARLIER FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/095,319
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: 60/095,454
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/095,455
; EARLIER FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 376
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 86
; LENGTH: 3174
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-489-847-86

Query Match   4.5%; Score 36.2; DB 4; Length 3174;
Best Local Similarity   52.3%; Pred. No. 0.43;
Matches 80; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```
Qy   224 GGAAGAGAGTAACACATTTTTTTTTTTTAATCCTGATAAAGAAGAATTGTTGGGAAGCTCTTTGA 283
Db  2285 GGAATACAATGGTTGTTGTTTTATCATAGTGTACACATTTAGCTTGTGTAAATGACTCACA 2344

Qy   284 AAAAAAATTTAAATTGTGGCACAGATGGGGATTTTAAAAAAGTGTTAGATCTTTCCAATGAA 343
Db  2345 AAACTGATTTTAAAATCAAGTTAATGTGAAATTTTGAAAAATTACTACTTAATCCTAATTCA 2404

Qy   344 CACTAATAGAGAGTACTCTGCTCCTTGGCTGGGATTT 376
Db  2405 CAATAACAATGGCCATTAAGGTTTGACTTGAGTT 2437
```

RESULT 13
US-08-545-528D-1/c
Sequence 1, Application US/08545528D
Patent No. 6537773
GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragments Thereof, and Uses Thereof
; FILE REFERENCE: PB193P1
; CURRENT APPLICATION NUMBER: US/08/545,528D
; CURRENT FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

Query Match   4.5%; Score 36.2; DB 4; Length 580073;
Best Local Similarity   56.2%; Pred. No. 8.8;
Matches 68; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy   199 TTTTTTTTTACTGTGTGGAAATAGGATCGGAGGAGAGTAACATTTTTTTTTAATCCTGATAA 258
Db 502669 TTATTTTTGCTGATCAAACAAAAAGAGAAGCAATTATTGGTGGTAATTTTTTTTTTAAATGTTATTA 502610

Qy   259 AGAAGATTGTTGGGAAGCTCTTTGAAAAAAAAATTTAAATTGTGGCACAGATGGGGATTTTA 318
Db 502609 AAAAAACAGAAGCTAAGATTTTCACAGAAAAATTTTGTTTTTGCTGATGAAGTTGGATATTA 502550

Qy   319 A 319
Db 502549 A 502549
```

RESULT 14
US-09-439-313-351/c
Sequence 351, Application US/09439313
Patent No. 6329905
GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang, Yuqui
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark

```
;   APPLICANT: Solk, John
;   APPLICANT: Day, Craig
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;   TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
;   FILE REFERENCE: 210121.427C9
;   CURRENT APPLICATION NUMBER: US/09/439,313
;   CURRENT FILING DATE: 1999-11-12
;   NUMBER OF SEQ ID NOS: 575
;   SOFTWARE: FastSEQ for Windows Version 3.0
;   SEQ ID NO 351
;   LENGTH: 472
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-439-313-351

Query Match             4.4%;  Score 35.8;  DB 4;  Length 472;
Best Local Similarity  51.6%;  Pred. No. 0.19;
Matches   79;  Conservative   1;  Mismatches   73;  Indels    0;  Gaps    0;

Qy    224 GGAAGAGAGTAACACATTTTTTTTTTTTTTAATCCTGATAAAGAAGAGATTGTTGGGAAGCTCTTTGA 283
                |||||          |||||      |||      ||      ||      ||
Db    206 GGAATACAAATGGTTGCTTTTTTATCATAGTGTTACACATTTAGCTTGTTGGGTAAATGACTCRCA 147

Qy    284 AAAAAAAAATTTTAAATTGTGGCACAGATGGGATTTTTAAAAAGTGTTAGATCTTTCCAAATGAA 343
                |:|   |||||||         ||||||   |||||||         ||||||
Db    146 AAWCTGATTTTAAAATCAAGTTAATGTGAATTTTGAAAAATTACTACTTAATCCTAATTCA 87

Qy    344 CACTAATAGAGTACTCTGCTCTTGGCTGGATTT 376
               |||||||      ||||||      |||||||
Db     86 CAATAACAAATGGCATTAAGGTTTGACTTGAGTT 54

RESULT 15
US-09-352-616A-351/c
Sequence 351, Application US/0935616A
Patent No. 6395278
GENERAL INFORMATION:
;   APPLICANT: Dillon, Davin C.
;   APPLICANT: Harlocker, Susan Louise
;   APPLICANT: Jiang, Yuqui
;   APPLICANT: Xu, Jiangchun
;   APPLICANT: Mitcham, Jennifer Lynn
;   TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;   TITLE OF INVENTION: OF PROSTATE CANCER AND METHODS FOR THEIR USE
;   FILE REFERENCE: 210121.427C8
;   CURRENT APPLICATION NUMBER: US/09/352,616A
;   CURRENT FILING DATE: 1999-07-13
;   NUMBER OF SEQ ID NOS: 472
;   SOFTWARE: FastSEQ for Windows Version 3.0
;   SEQ ID NO 351
;   LENGTH: 472
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-352-616A-351

Query Match             4.4%;  Score 35.8;  DB 4;  Length 472;
Best Local Similarity  51.6%;  Pred. No. 0.19;
Matches   79;  Conservative   1;  Mismatches   73;  Indels    0;  Gaps    0;

Qy    224 GGAAGAGAGTAACACATTTTTTTTTTTTTTAATCCTGATAAAGAAGAGATTGTTGGGAAGCTCTTTGA 283
                |||||          |||||      |||      ||      ||      ||
Db    206 GGAATACAAATGGTTGCTTTTTTATCATAGTGTTACACATTTAGCTTGTTGGGTAAATGACTCRCA 147

Qy    284 AAAAAAAAATTTTAAATTGTGGCACAGATGGGATTTTTAAAAAGTGTTAGATCTTTCCAAATGAA 343
                |:|   |||||||         ||||||   |||||||         ||||||
Db    146 AAWCTGATTTTAAAATCAAGTTAATGTGAATTTTGAAAAATTACTACTTAATCCTAATTCA 87

Qy    344 CACTAATAGAGTACTCTGCTCTTGGCTGGATTT 376
               |||||||      ||||||      |||||||
Db     86 CAATAACAAATGGCATTAAGGTTTGACTTGAGTT 54
```

Search completed: April 9, 2004, 15:46:45

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: April 9, 2004, 15:02:33 ; Search time 355 Seconds
                 (without alignments)
                 8580.808 Million cell updates/sec

Title: US-09-763-335-4
Perfect score: 812
Sequence: 1 cactggagtggggatggtccc..........catttggtagtaaaggaa 812

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 2475585 seqs, 1875730760 residues

Total number of hits satisfying chosen parameters: 4951170

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Database :   Published_Applications_NA:*
           1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
           2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
           3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
           4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
           5: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
           6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
           7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
           8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
           9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
          10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
          11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
          12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
          13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
          14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
          15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
          16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
          17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
          18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	131.2	16.2	716	14	US-10-037-270-225	Sequence 225, App
2	131.2	16.2	716	15	US-10-117-722-225	Sequence 225, App
3	128.4	15.8	2390	15	US-10-104-047-892	Sequence 892, Appl
4	43.4	5.3	3673778	14	US-10-312-841-1	Sequence 1, Appli
5	42.6	5.2	16724	14	US-10-311-455-1064	Sequence 1064, Ap
6	42.6	5.2	16724	14	US-10-240-485-90	Sequence 90, Appl
7	42	5.2	498	10	US-09-991-936-1045	Sequence 1045, Ap
8	41.8	5.1	10886	10	US-10-311-455-2108	Sequence 2108, Ap
9	41.4	5.1	40862	14	US-10-311-455-2046	Sequence 2046, Ap
10	41	5.0	17869	16	US-10-311-455-78	Sequence 78, Appl
11	41	5.0	17869	16	US-10-257-166-2	Sequence 2, Appl
12	40.4	5.0	13511	14	US-10-311-455-253	Sequence 253, Appl
13	40.2	5.0	9268	14	US-10-311-455-1573	Sequence 1573, Ap
14	40.2	5.0	3673778	14	US-10-312-841-2	Sequence 2, Appli
15	39.8	4.9	13038	14	US-10-311-455-1248	Sequence 1248, Ap
16	39.6	4.9	5875	14	US-10-311-455-262	Sequence 262, App
17	39.6	4.9	11416	12	US-10-221-613-19	Sequence 19, Appl
18	39.6	4.9	11416	14	US-10-221-613-91	Sequence 91, App
19	39.4	4.9	89328	10	US-09-873-367C-332	Sequence 332, App
20	39.2	4.8	6904	14	US-10-311-455-186	Sequence 186, App
21	39.2	4.8	23934	9	US-09-764-877-2455	Sequence 2455, Ap
22	39.2	4.8	23934	15	US-10-242-515-2455	Sequence 2455, Ap
23	39	4.8	7276	14	US-10-311-455-875	Sequence 875, App
24	38.8	4.8	8979	12	US-10-221-613-138	Sequence 138, App
25	38.8	4.8	8979	14	US-10-311-455-758	Sequence 758, App
26	38.6	4.8	5269	14	US-10-311-455-2029	Sequence 2029, Ap
27	38.2	4.7	652	15	US-10-125-968-1355	Sequence 1355, Ap
28	38.2	4.7	16914	14	US-10-221-613-214	Sequence 214, App
29	38	4.7	10716	14	US-10-311-455-1391	Sequence 1391, Ap
30	37.8	4.7	564	15	US-10-027-632-286621	Sequence 286621,
31	37.8	4.7	564	15	US-10-027-632-286628	Sequence 286628,
32	37.8	4.7	564	15	US-10-027-632-286629	Sequence 286629,
33	37.8	4.7	564	15	US-10-027-632-286630	Sequence 286630,
34	37.6	4.7	14253	14	US-10-027-632-1467	Sequence 1467, Ap
35	37.4	4.6	5970	14	US-10-108-605-210	Sequence 210, App
36	37.4	4.6	6282	13	US-10-108-605-212	Sequence 212, App
37	37.4	4.6	7049	14	US-10-240-452-5	Sequence 129, App
38	37.4	4.6	7049	14	US-10-240-452-5	Sequence 5, Appli
39	37.4	4.6	9155	14	US-10-311-455-435	Sequence 435, App
40	37.4	4.6	17848	14	US-10-239-677-28	Sequence 28, Appl
41	37.4	4.6	17848	14	US-10-240-453-38	Sequence 38, Appl
42	37.2	4.6	17848	16	US-10-257-166-58	Sequence 58, Appl
43	37.2	4.6	8136	14	US-10-311-455-527	Sequence 58, Appl
44	37.2	4.6	8136	16	US-10-257-166-37	Sequence 37, Appl
45	37	4.6	532	10	US-09-814-353-13846	Sequence 13846, A

ALIGNMENTS

RESULT 1
US-10-037-270-225
Sequence 225, Application US/10037270
Publication No. US20030104529A1

GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yunqing
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: Tillinghast, John
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. US20030104529A1e1 Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/10/037,270
CURRENT FILING DATE: 2002-01-04
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1104
SOFTWARE: pt_FL_genes Version 1.0
SEQ ID NO 225
LENGTH: 716
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:

; NAME/KEY: CDS
; LOCATION: (201)...(596)
US-10-037-270-225

Query Match   16.2%;   Score 131.2;   DB 14;   Length 716;
Best Local Similarity   63.1%;   Pred. No. 3.6e-27;
Matches 202; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

```
Qy  479 AGCAGCATCACCTGCACAGACCAGGAGGAGGACGTGTGAAGTGATAGCAGCACCGGAT 538
        ||| ||  ||    |      ||      |||||||||
Db  295 ACCATCATAAAGCTCACCATGTTAAAACGGGGAACTTGTGAGGTGGTGGCACTCCACAGAT 354

Qy  539 GTTGTAACAAGAATCGCATTGGAGGAGGGTCACAAACAGTAAAGTGTTCCTGTTCTACCTG 598
        |||||||||||||||||||||||||||||||
Db  355 GCTGTAATAAGAACAAGATAGAAGAACAGGTCACAAAACAGTCAAGTGCTTCAATAGTGGAAC 414

Qy  599 GAAAAGTGGCTGGAACAACAAGAAAACCGGGTCACAAACAGTAAAGTGTTCCTGTTCTACCTG 658
        ||  |||||||
Db  415 GGCAGGTGGCAGGCACCACGCGAGCTGCTCCATCATGTGTGGATGCTTCAATAGTGGAAC 474

Qy  659 GGAAATGGTGGTGGTGTGAGATGGAGCCTTGCCTAGAGAGGAGGAATGTAAGACACTCCCTG 718
        ||||||||||
Db  475 AGAAATGGTGGTGCGCATAGCAGCCATGTCTTAGACGGGAGAGAAATGTAAAGTTCTTCCGG 534

Qy  719 ACAATTCTGGATGGATGCGGCAACAGGCACCCAACACACGAGAAATTCACCCAA 778
        ||||||||
Db  535 ATCGGAAAGGATGGAGCTGTTCCTCTGGGAATAAAGTCAAAAACAACTAGGGTAACCCATT 594

Qy  779 GAACCTAACAGAAGCATTTG 798
        ||  ||
Db  595 AACCCAGGAGAGAAATCAAGTG 614
```

RESULT 2
US-10-117-722-225
; Sequence 225, Application US/10117722
; Publication No. US20030219744A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2BCIP
; CURRENT APPLICATION NUMBER: US/10/117,722
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 225
; LENGTH: 716
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)...(596)
US-10-117-722-225

Query Match   16.2%;   Score 131.2;   DB 15;   Length 716;
Best Local Similarity   63.1%;   Pred. No. 3.6e-27;
Matches 202; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

```
Qy  479 AGCAGCATCACCTGCACAGACCAGGAGGAGGACGTGTGAAGTGATAGCAGCACCGGAT 538
Db  295 ACCATCATAAAGCTCACCATGTTAAAACGGGGAACTTGTGAGGTGGTGGCACTCCACAGAT 354
```

RESULT 3
US-10-104-047-892
; Sequence 892, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 892
; LENGTH: 2390
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-892

Query Match   15.8%;   Score 128.4;   DB 15;   Length 2390;
Best Local Similarity   92.5%;   Pred. No. 4.6e-26;
Matches 135; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy  357 CTCTGCTCTTGGCCTGGATTTTTCAGAGAATGGCAATGGTTCTCTGCGATGGTCCTGGGGTCCT 416
        |||| ||| || ||||||||||||||||||||||||||||
Db  610 CTTTCCTGTCGGAATGTTCTTCTTTAGAGAATGGGCAATGGTTCTCTGCGATGTCCTGGGGTCCT 669

Qy  417 GTATTTGTGGATAAGTGCTTGTTGTGCAATGCTACTTACTCCTGCCATGAATCCCTTCACACACTTT 476
        ||||||||||||||||||||||||
Db  670 GTATTTGTGGATAAGTGCTTGTTGTGCAATGCTACTCTGCAATGGCCATGATCCCTTCAGACACTTT 729

Qy  477 CCAGCAGCATCACCTGCACAGACCAG 502
        ||||||||||||||||||||
Db  730 CCAGCAGCATCACCTGCACAGACCAG 755
```

RESULT 4
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial sequence

```
Qy  479 AGCAGCATCACCTGCACAGACCAGGAGGAGGACGTGTGAAGTGATAGCAGCACCGGAT 538
Db  295 ACCATCATAAAGCTCACCATGTTAAAACGGGGAACTTGTGAGGTGGTGGCACTCCACAGAT 354
```

```
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;     NAME/KEY: unsure
;     LOCATION: (3294164)
US-10-312-841-1

Qy    239  TTTTTTTTTTAATCCTGATAAAGAAGAANTGTTGGGGAAGCTCCTTGAAAAAAATTTTAAAT  298
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2721915  TGTTTTTTTTTTATTTTTTATTGAGTAGTAGTTGTTGTAGAATTTTTTTATAT  2721974

Qy    299  TGTGCACAGATGGAGATTTAAAAAGTGTTAGATCTTTCCAATGAACACTAATAGAGTACT  358
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2721975  TTTTTCTTAAGTTGGGATTTTATTTTTTTGAAGTAATTGTGAATGGGAGTT  2722034

Qy    359  CTGCCTCTTGGCGTGGATTTTTCAGAGAANTGGCAANTGGCTCCTGCGATGTCCTGGGTCCTGT  418
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2722035  TATTTATGATTTGGTTTTGTTTTGTTATTGGTGTATAAGAANTGTTTGTTGATTTTT  2722094

Qy    419  ATTTGTGGATAAGTGCTTGTGCAAT  443
           ||||||||||||||||||||||||
Db 2722095  GTATATATTGATTTTGTATTTGAGAT  2722119

RESULT 5
US-10-311-455-1064
;   Sequence 1064, Application US/10311455
;   Publication No. US20030143606A1
;   GENERAL INFORMATION:
;   APPLICANT: OLEK, Alexander
;   APPLICANT: PIEPENBROCK, Christian
;   APPLICANT: BERLIN, Kurt
;   TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;   TITLE OF INVENTION: cytosine methylation
;   FILE REFERENCE: 5013.1014
;   CURRENT APPLICATION NUMBER: US/10/311,455
;   CURRENT FILING DATE: 2002-12-16
;   PRIOR APPLICATION NUMBER: PCT/EP01/07537
;   PRIOR FILING DATE: 2001-07-02
;   PRIOR APPLICATION NUMBER: DE 10032529.7
;   PRIOR FILING DATE: 2000-06-30
;   PRIOR APPLICATION NUMBER: DE 10043826.1
;   PRIOR FILING DATE: 2000-09-01
;   NUMBER OF SEQ ID NOS: 2424
SEQ ID NO: 1064
;   LENGTH: 16724
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1064

Query Match    5.2%;  Score 42.6;  DB 14;  Length 16724;
Best Local Similarity   46.2%;  Pred. No. 0.85;
Matches  141;  Conservative    0;  Mismatches  164;  Indels    0;  Gaps    0;

Qy    126  AAAATGACAAGGTTTCCACCCCCTCAAACCTTGGCTCCTTTTCTGACAATACAGTCTGAAT  185
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13695  AAGTTGAGATGGTGTTATTTGGGTTATTAGAGTTGAGATTTTATTTTTTTAAAA  13754

Qy    186  GAACCCGANTGTCTTTTTTTTTACTGTGGAAATAGGATCGGAAGAGAGTAACATTTTTTT  245
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13755  AAAAAATGTATTAAATATTTTTATTTTGATTAAATTTAAAGATATGTAGTTAAAATTT  13814

Qy    246  TTAATCCTGATAAAGAAGADATTGTTGGGGAAGCTCCTTTTGAAAAAAAATTTTAAAATGTGGCA  305
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13815  TGGTATGTTTTATTAAGGTATAGGGTTTAGTAATTTGTTTGAAAGAAAAATTATTGATTGAGAGAA  13874

Qy    306  CAGATGGATTTTAAAAAGTGTTAGATCTTTCCAATGAACACTAATAGAGTACTCTGCTCT  365
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13815  TGGTATGTTTTATTAAGGTTTAGATCTTGTATTTTGTTTGAAAGAAAAATTATTGATTGAGAGA  13874

Qy    246  TTAATCCTGATAAAGAAGADATTGTTGGGGAAGCTCCTTTTGAAAAAAAATTTTAAAATGTGGCA  305
Db 13815  TGGTATGTTTTATTAAGGTTTAGATCTTGTATTTTGTTTGAAAGAAAAATTATTGATTGAGAGA  13874

Qy    306  CAGATGGATTTTAAAAAGTTTAGATCTTTCAATGAACACTAATAGAGTACTCTGCTCT  365

RESULT 6
US-10-240-485-90
;   Sequence 90, Application US/10240485
;   Publication No. US20030148327A1
;   GENERAL INFORMATION:
;   APPLICANT: OLEK, Alexander
;   APPLICANT: PIEPENBROCK, Christian
;   APPLICANT: BERLIN, Kurt
;   TITLE OF INVENTION: Diagnosis of Diseases Associated with
;   TITLE OF INVENTION: Metastasis
;   FILE REFERENCE: 5013.1007
;   CURRENT APPLICATION NUMBER: US/10/240,485
;   CURRENT FILING DATE: 2002-10-02
;   PRIOR APPLICATION NUMBER: PCT/EP01/03970
;   PRIOR FILING DATE: 2001-04-06
;   PRIOR APPLICATION NUMBER: DE 10019058.8
;   PRIOR FILING DATE: 2000-04-06
;   PRIOR APPLICATION NUMBER: DE 10019173.8
;   PRIOR FILING DATE: 2000-04-07
;   PRIOR APPLICATION NUMBER: DE 10032529.7
;   PRIOR FILING DATE: 2000-06-30
;   PRIOR APPLICATION NUMBER: DE 10043826.1
;   PRIOR FILING DATE: 2000-09-01
;   NUMBER OF SEQ ID NOS: 202
SEQ ID NO: 90
;   LENGTH: 16724
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-90

Query Match    5.2%;  Score 42.6;  DB 14;  Length 16724;
Best Local Similarity   46.2%;  Pred. No. 0.85;
Matches  141;  Conservative    0;  Mismatches  164;  Indels    0;  Gaps    0;

Qy    126  AAAATGACAAGGTTTCCACCCCCTCAAACCTTGGCTCCTTTTCTGACAATACAGTCTGAAT  185
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13695  AAGTTGAGATGGTGTTATTTGGGTTATTAGAGTTGAGATTTTATTTTTTTAAAA  13754

Qy    186  GAACCCGATGTCTTTTTTTTTTACTGTGGAAATAGGATCGGAAGAGAGTAACATTTTTTTT  245
           |||||||||||||||||||||||||||||||||||||||||
Db 13755  TTAATATATTAAAGAGGTTATTGATGATGTGATAAAATTTTTTTTATTTT  13814

Qy    246  TTAATCCTGATAAAGAAGATTGTTGGGGAAGCTCCTTTTGAAAAAAAATTTTAAAATGTGGCA  305
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13815  TGGTATGTTTTATTAAGGTTTAGATCTTGTATTTTGTTTGAAAGAAAAATTATTGATTGAGAGAA  13874

Qy    306  CAGATGGATTTTAAAAAGTGTTAGATCTTTCCAATGAACACTAATAGAGTACTCTGCTCT  365
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13875  GAAAATTTATATAAAATTTATATAAAATGTAAAATAGTTTATTTAATTTTTATTTTTTTT  13934

Qy    366  TGGCTGGATTTTTCAGAGAATGGCAAATGGTCTCTCGCGATGTCCTGGGTCCTGTATTTGTG  425
           ||||||||||||||||||||||||||||||||||||||||
Db 13935  TTAATATATTAAAAGGAGGTTATTGATATGTATTGTATAAAAATTAAAGGGGTTTTTTTATATTTT  13994

Qy    426  GATAA  430
           ||||
Db 13995  AAAAA  13999
```

RESULT 7
US-09-991-936-1045/c
; Sequence 1045, Application US/0991936
; Publication No. US20030073827A1
; GENERAL INFORMATION:
; APPLICANT: Brandt, Kevin S.
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Wisnewski, Nancy
; TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES, PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-6-C1
; CURRENT APPLICATION NUMBER: US/09/991,936
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US/09/543,668
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,704
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 1959
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1045
; LENGTH: 498
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-991-936-1045

  Query Match   5.2%; Score 42; DB 10; Length 498;
  Best Local Similarity   47.9%; Conservative 0; Mismatches 114; Indels 0; Gaps 0;
  Matches 105;

Qy   164 TTTCTGACAATACAGTCTGAATGAACCCGATGTCTTTTTTTTTTACTGTGGAAATAGGATC 223
          ||||||| ||| | |||| |||||| | | ||||||||||||||| |||| |||||||||
Db   486 TTTTTTAANTGANAGGTGGBAATGAATGAATGAAAATATTAACTGTCTTAATANAATT 427

Qy   224 GGAGAGAGTAACATTTTTTTTTTTTAATCCTGATAAAGAAGAGAGATTGTTGGGAAGCTCTTTGA 283
          ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   426 GTTATAGAATNTAATTTTTAGTGTTAAANAGGGTAAATAATTATTTTAAAGGCGAGAAGACCCT 367

Qy   284 AAAAAAATTTTAAATGTGGCACAGATGGATTTTAAAAAGTGTTAGATCTTTCCAATGAA 343
          ||||||| |||||||| || |||||||||||||||| |||||||||||||||||||||||
Db   366 TATAAAGTTTTATATATNATATNAATTTTAAGAATATATTATTATANANTTTATTTATA 307

Qy   344 CACTAATAGAGTACTCTGCTCTTTGGCTGCTGATTTTCAGA 382
          ||||||||||||||||| |||||| || ||||||
Db   306 TTAATACATTTATTGGGCGTGAATAGGAANAATTATTATAAAA 268

RESULT 8
US-10-311-455-2108
; Sequence 2108, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2108
; LENGTH: 10886
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-311-455-2108

  Query Match   5.1%; Score 41.8; DB 14; Length 10886;
  Best Local Similarity   50.2%; Pred. No. 1.1;
  Matches 103; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

Qy   239 TTTTTTTTTTAATCCTGATAAAGAAGAGAGATTGTTGGGAAGCTCTTTGAAAAAAATTTTAAAT 298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db 10668 TGTTTTTTTTTTATTTATTGAGTAGTGCGTTCGTAGTTTTTTTTTTGAAGTAGTTTTTACGT 10727

Qy   299 TGTGGCACAGATGGATTTTAAAAAGTGTTAGATCTTTCCAATGAACACTAATAGAGTACT 358
          |||| | |||||||||| ||| ||| | ||| ||||||| |||||||| ||||||||
Db 10728 TTTTGTAAGTTGGGATTTTTAAATTAAGTAGTATTTTTTATTTTTTGAAGTAGATAATTAGAGAGT 10787

Qy   359 CTGCTCTTTGGCTGGATTTTTTCAGAGATGGCAAATGGTTCTCTCGGATGTCTCTTGA 418
          ||||||||||| |||| | || ||| || |||||||||||||||||||||||||
Db 10788 TATTTATGATTTGGTTTTTTTTGTTTTTTGTTTTGTTTATTGGGTGTATAAGGAATGTTTGTGATTTTT 10847

Qy   419 ATTTGTGGATAAGTAGCTTGTGCAAT 443
          |||||||| |||| ||||| ||||||
Db 10848 GTATATGATTTGTTATTTGAGAT 10872

RESULT 9
US-10-311-455-2046
; Sequence 2046, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2046
; LENGTH: 40862
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2046

  Query Match   5.1%; Score 41.4; DB 14; Length 40862;
  Best Local Similarity   57.3%; Pred. No. 3.2;
  Matches 75; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Qy   193 ATGTCTTTTTTTTTTACTGTGGAAATAGGATCCGGAGAGAGTAACATTTTTTTTTTTTAATCC 252
          || | |||||||||||||||||||||||| ||||| |||||||| |||||||||
Db   1414 ATGTTTTTTATTTTTAAATTTATAAAAATTAAAATTAGGATAGGGAAGAAAGAGATTTTATTGAAATCGT 1473

Qy   253 TGATAAAGAAGAGAGATTGTTGGGAAGCTCTTTGAAAAAAAAATTTTAAATTGTGGCACAGATGG 312
          ||| |||| | ||| | ||||||||||| || |||||||||||||||| |||
Db   1474 TTTTAAGTAGTAAAAATATTGGAGAGAGATTTTTTTTTTAAAATGATAAATAAATAATGAG 1533

Qy   313 ATTTTAAAAAG 323
          ||||| ||| ||
Db   1534 AATTGAAATAG 1544

RESULT 10
US-10-311-455-78
; Sequence 78, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:

```
;   APPLICANT: OLEK, Alexander
;   APPLICANT: PIEPENBROCK, Christian
;   APPLICANT: BERLIN, Kurt
;   TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;   TITLE OF INVENTION: Cytosine methylation
;   FILE REFERENCE: 5013.1014
;   CURRENT APPLICATION NUMBER: US/10/311,455
;   CURRENT FILING DATE: 2002-12-16
;   PRIOR APPLICATION NUMBER: PCT/EP01/07537
;   PRIOR FILING DATE: 2001-07-02
;   PRIOR APPLICATION NUMBER: DE 10032529.7
;   PRIOR FILING DATE: 2000-06-30
;   PRIOR APPLICATION NUMBER: DE 10043826.1
;   PRIOR FILING DATE: 2000-09-01
;   NUMBER OF SEQ ID NOS: 2424
;   SEQ ID NO 78
;   LENGTH: 17869
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 3465, 3586
;   OTHER INFORMATION: n is a or g or c or t
US-10-311-455-78

Query Match            5.0%;   Score 41;  DB 14;  Length 17869;
Best Local Similarity  54.2%;   Pred. No. 2.5;
Matches   83;  Conservative   0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy       200 TTTTTTTACTGTGGAAATAGGATCGGAAGAGGAGAGTAACACATTTTTTTTTTTTAATCCTGATAAA 259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14425 TTTTTTTTAGTGAATAGTTATTTTTTATAGGTATTTAAATAATTTTTTTTTTTTTTTATTTAAGA 14484

Qy       260 GAAGATTGTTGGGAAGCTCTTTGAAAAAAAAAATTTAAATTGTGGCACAGATGGATTTTAA 319
              ||||||||||||||||||    ||||||||||||   |||||||| |||||
Db     14485 TAAGAATTTGTAAAAATTTTTTTTTGTAAATATTTTAGTTTTTGGGTGGATGGTTTTTTGA 14544

Qy       320 AAAGTGTTAGATCTTTCCAATGAACACTAATAG 352
              |||||||  |  ||   |  |   ||  |  |||
Db     14545 TATAATTTATTTGATTTTATTTAAAAGAAATTG 14577
```

```
US-10-257-166-2
Sequence 2, Application US/10257166
Publication No. US20040023230A1
GENERAL INFORMATION:
;   APPLICANT: OLEK, Alexander
;   APPLICANT: PIEPENBROCK, Christian
;   APPLICANT: BERLIN, Kurt
;   TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
;   TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
;   FILE REFERENCE: 5013.1011
;   CURRENT APPLICATION NUMBER: US/10/257,166
;   CURRENT FILING DATE: 2002-10-07
;   PRIOR APPLICATION NUMBER: PCT/EP01/07470
;   DE 10032529.7
;   PRIOR FILING DATE: 2001-06-29
;   DE 10043826.1
;   2000-06-30
;   2000-09-01
;   NUMBER OF SEQ ID NOS: 178
SEQ ID NO 2
;   LENGTH: 17869
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3465, 3586)
```

```
US-10-257-166-2

Query Match            5.0%;   Score 41;  DB 16;  Length 17869;
Best Local Similarity  54.2%;   Pred. No. 2.5;
Matches   83;  Conservative   0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy       200 TTTTTTTACTGTGGAAATAGGATCGGAAGAGGAGAGTAACACATTTTTTTTTTTTAATCCTGATAAA 259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14425 TTTTTTTTAGTGAATAGTTATTTTTTATAGGTATTTAAATAATTTTTTTTTTTTTTTATTTAAGA 14484

Qy       260 GAAGATTGTTGGGAAGCTCTTTGAAAAAAAAAATTTAAATTGTGGCACAGATGGATTTTAA 319
              ||||||||||||||||||    ||||||||||||   |||||||| |||||
Db     14485 TAAGAATTTGTAAAAATTTTTTTTTGTAAATATTTTAGTTTTTGGGTGGATGGTTTTTTGA 14544

Qy       320 AAAGTGTTAGATCTTTCCAATGAACACTAATAG 352
              |||||||  |  ||   |  |   ||  |  |||
Db     14545 TATAATTTATTTGATTTTATTTAAAAGAAATTG 14577
```

RESULT 12
```
US-10-311-455-253
Sequence 253, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
;   APPLICANT: OLEK, Alexander
;   APPLICANT: PIEPENBROCK, Christian
;   APPLICANT: BERLIN, Kurt
;   TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;   TITLE OF INVENTION: Cytosine methylation
;   FILE REFERENCE: 5013.1014
;   CURRENT APPLICATION NUMBER: US/10/311,455
;   CURRENT FILING DATE: 2002-12-16
;   PRIOR APPLICATION NUMBER: PCT/EP01/07537
;   PRIOR FILING DATE: 2001-07-02
;   PRIOR APPLICATION NUMBER: DE 10032529.7
;   PRIOR FILING DATE: 2000-06-30
;   PRIOR APPLICATION NUMBER: DE 10043826.1
;   PRIOR FILING DATE: 2000-09-01
;   NUMBER OF SEQ ID NOS: 2424
;   SEQ ID NO 253
;   LENGTH: 13511
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 4623
;   OTHER INFORMATION: n is a or g or c or t
US-10-311-455-253

Query Match            5.0%;   Score 40.4;  DB 14;  Length 13511;
Best Local Similarity  50.5%;   Pred. No. 3.2;
Matches   98;  Conservative   0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy       158 GCTCCTTTCTGACAATACAGTCTGAATGAACCCGATGTGTCTTTTTTTTTTTACTGTGGAAAT 217
              |||  |||||  |||||    ||||| ||||||||||||||||||||||||||||
Db     12024 GATTATATTAATTGTATTTGATTTTGGGTGATAGAGTGAGAGATTTATTTTAAAAAAAAAAA 12083

Qy       218 AGGATCGGAAGAGAGTAACATTTTTTTTTTTTTAATCCCTGATAAAGAAGATTGTTGGGAAGCT 277
              |||||| |||||||||||   ||||||||||||||||||||    |||||||||
Db     12084 AAAAAAGAAAAAAAAAAGTTTATTGATTTTATTTAATAAGAAAATATTTTAGTTTTTGGGAATA 12143

Qy       278 CTTTGAAAAAAAAATTTAAATGTGGCACAGATGGATTTTTAAAAAAGTTGTTAGATCTTTCC 337
              ||||||||||  ||||||  ||| |||||||||||||   |||||   || ||  ||
Db     12144 TATTGAAAATAAATAATTTTTTATTTTAGTTTTTAAGTTTGTTAGGATTTAGA 12203

Qy       338 AATGAACACTAATA 351
              ||||||   || ||
Db     12204 AATGATGAATTTTA 12217
```

RESULT 13

US-10-311-455-1573
; Sequence 1573, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1573
; LENGTH: 9268
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

Query Match   5.0%; Score 40.2; DB 14; Length 9268;
Best Local Similarity 47.8%; Pred. No. 3; Mismatches 128; Indels 0; Gaps 0;
Matches 117; Conservative 0;

Qy   179 TCTGAATGAACCGATGTCTCTTTTTTTTTTTACTGTGGAAATAGGATCGGAAGAGAGAGTAACACAT 238
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1661 TTTGTTTGGGATGGATGAAGTTTTTTTTTTTTAGAGTAGTTTTTGGTTTGTTTTGGGAGAGTTTATG 1720

Qy   239 TTTTTTTTTTTAATCCTGATAAAGAAGAAGATTGTTGGGGAAGCTCTTTGAAAAAAAATTTTAAAT 298
               |||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1721 GTTTTTAGTTTAGATTTAATAAATCGATTTATATATATATTTTAATATATTTTAATTGTTTTGGT 1780

Qy   299 TGTGGCACAGATGGATTTTAAAAAGTGTTAGATCTTTCCAATGAACACTAATAGAGTACT 358
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1781 TAAGGAGTATAAAGTTTTTTTTTTTTTTGATAGAAGATGTTTATGTGTTTTTTAAAATAAT 1840

Qy   359 CTGCTCTTGGCTCGGATTTTTCAGAGAATGGCAAATGGGTCTCTCGGATGTCCTGGGTCCTGT 418
               ||| ||||| ||| |||||||
Db   1841 TAGGTTATGATTTTTTTATTTTTAGTATATTAATATAGGAAATATTTGTGATTTTGTGTTTTTT 1900

Qy   419 ATTTG 423
               |||||
Db   1901 TTTGG 1905

RESULT 14
US-10-312-841-2
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (379615)
US-10-312-841-2

Query Match   5.0%; Score 40.2; DB 14; Length 3673778;
Best Local Similarity 47.8%; Pred. No. 93; Mismatches 128; Indels 0; Gaps 0;
Matches 117; Conservative 0;

Qy   199 TTTTTTTTTTACTGTGGAAATAGGATCGGAAGAGAGAGTAACACATTTTTTTTTTTTAATCCTGATAA 258
               ||||| |||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db 2860776 TTTTTTTTTATTTATGAGTATAGAAATGTTTTTCGTTTGTTTTGGCACGATGGGATTTTA 2860835

Qy   259 AGAAGAATTGTTGGGGAAGCTCTTTGAAAAAAAATTTTAAAATGTGGCACGATGGGATTTTA 318
               |||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db 2860836 AGTAGTGGGTTTGTAGTTTTTGGTTTTTTTGAAGAGGTTTTTTATATTTTTTTGTTTGGT 2860895

Qy   319 AAAAAGTGTTAGATCTTTCCAATGAACACTAATAGAGTACTCTGCTCTTGGCTCGGATTTTT 378
               |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2860896 AGGTATTTTTATTTTTTGTAGTAATGTGAATGGGAGTTTATTTATTTAATGATTTGGTTTTTTT 2860955

Qy   379 CAGAGAATGGCAAATGGGTCTCTCGGATGTCCTGGGTCCTGTATTTGTGGATAAGTAGTTGCTTGT 438
               |||| |||||| || ||||||| ||| ||| ||| ||||||||
Db 2860956 GTTTGTTTGTTATTGGTATATAGGGAAATATTTGTGATTTTGTGTTTTTTTTTGGGGGTTTTTT 2861015

Qy   439 GCAAT 443
               |||||
Db 2861016 GAGAT 2861020

RESULT 15
US-10-311-455-1248
; Sequence 1248, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1248
; LENGTH: 13038
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1248

Query Match   4.9%; Score 39.8; DB 14; Length 13038;
Best Local Similarity 57.7%; Pred. No. 4.7; Mismatches 52; Indels 0; Gaps 0;
Matches 71; Conservative 0;

Qy   234 AACATTTTTTTTTTTTTAATCCTGATAAAGAAGAAGATTGTTGGGGAAGCTCTTTGAAAAAAAATTT 293
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9690 AATATGTGTTTTAGAGAAGAAGAAAAAGAAAATTATGAATAGAATTAATTTAATTTAAATTG 9749

Qy   294 TAAAATTGTGGCACAGATGGATTTTAAAAAGTGTTAGATCTTTCCAATGAACACTAATAGA 353
               |||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   9750 TAAGTTGTTTGAAGATTGTATTTTTAAAAAATGTTATTTGTTATTAAGTTGTTAAAAAAAA 9809

Qy   354 GTA 356
               |||
Db   9810 GAA 9812

Search completed: April 9, 2004, 16:51:23
Job time : 373 secs

```
OM nucleic - nucleic search, using sw model

                        GenCore version 5.1.6
                        Copyright (c) 1993 - 2004 Compugen Ltd.

Run on:         April  9, 2004, 12:42:37 ; Search time 2543 Seconds
                                            (without alignments)
                                            9535.230 Million cell updates/sec

Title:          US-09-763-335-4
Perfect score:  812
Sequence:       1 cactggagtggggatggtgtcc............cattcgtggtagtaaaggaa 812
Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters:        55026578

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      EST:*
                1: em_estba:*
                2: em_esthum:*
                3: em_estin:*
                4: em_estmu:*
                5: em_estov:*
                6: em_estpl:*
                7: em_estro:*
                8: em_htc:*
                9: gb_est1:*
               10: gb_est2:*
               11: gb_htc:*
               12: gb_est3:*
               13: gb_est4:*
               14: gb_est5:*
               15: em_estfun:*
               16: em_estom:*
               17: em_estfun:*
               18: em_gss_hum:*
               19: em_gss_inv:*
               20: em_gss_pln:*
               21: em_gss_vrt:*
               22: em_gss_fun:*
               23: em_gss_mam:*
               24: em_gss_mus:*
               25: em_gss_pro:*
               26: em_gss_rod:*
               27: em_gss_phg:*
               28: em_gss_vrl:*
               29: gb_gss1:*
               29: gb_gss2:*

                                SUMMARIES

Result                    %       Query
No.    Score  Match  Length  DB ID       Description
-----  -----  -----  ------  --------    -----------
1      812    100.0   1934    11  BC025746    BC025746 Homo sapi
2      800.2   98.5    929     9  AL534074    AL534074 AL534074
3      767.2   94.5    940    13  BX460561    BX460561 BX460561
4      667     82.1    793    12  BI756155    BI756155 603030059
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```
    5   610.2    75.1   1022  12  BI488689    BI488689 603021235
    6   551.6    67.9   3258  11  AK050001    AK050001 Mus muscu
    7    511     62.9    672  14  CK001752    CK001752 AGENCOURT
    8   475.4    58.5    768  13  BU118807    BU118807 603143059
    9   442.2    54.5    506  10  BF652347    BF652347 275877 MA
   10   440.8    54.3    706  13  BU615139    BU615139 UI-M-FRO-
   11   434.8    53.5    975   9  AL533965    AL533965 AL533965
   12    409     50.4    468  14  H23443      H23443 ym52g09.r1
   13   389.2    47.9    692  10  BB653506    BB653506 BB653506
   14   366.6    45.1   2412  11  AK049880    AK049880 Mus muscu
   15   361.8    44.6    845  13  BU275101    BU275101 603532890
   16   353.4    43.5    889  13  BU272791    BU272791 603531380
   17    349     43.0    728  10  AW955725    AW955725 EST367795
   18   346.8    42.7    443  14  R67171      R67171 yh08d07.r2
   19   320.4    39.5    442  14  R56299      R56299 yg90b04.r1
   20   318.4    39.2    853  14  T81284      T81284 yd06g02.r1
   21    294     36.2    357  13  BU281041    BU281041 603866055
   22    276     34.0    472  28  B2950260    B2950260 CH240_390
   23   274.6    33.8    473  14  CB730363    CB730363 AMGNNUC:N
   24   257.8    31.7    251   9  BB647606    BB647606 UI-M-BH1-
   25    251     30.9    420  14  AA365782    AA365782 EST76613
   26    215     26.5    594  10  T80333      T80333 ydz07d07.r1
   27    201     24.8    558  10  BB653101    BB653101 BB653101
   28    200     24.6    510   9  AL861103    AL861103 AL861103
   29   199.4    24.6    643   9  AL861104    AL861104 AL861104
   30   199.4    24.6    406  13  BY283942    BY283942 BY283942
   31   197.8    24.4    451  13  BY256413    BY256413 BY256413
   32   187.6    23.1    484   9  AW049606    AW049606 UI-M-BH1-
   33   181.6    22.4    436  13  BY283473    BY283473 BY283473
   34   173.8    21.4    894  13  BU133938    BU133938 BU133938
   35   173.8    21.4    460  14  R13004      R13004 yf71b07.r1
   36   164.2    20.3    538  28  AQ318831    AQ318831 RPCIII-10
   37    145     17.9    440  13  BY250323    BY250323 BY250323
   38   144.2    17.8    778  13  BX101385    BX101385 BX101385
   39   143.2    17.6    535  29  CG561270    CG561270 OST182659
   40   139.4    17.2    397  29  AY404310    AY404310 Homo sapi
   41   136.8    16.8    397  29  AY404312    AY404312 Homo muscu
   42   136.8    16.8    637  13  BX085931    BX085931 BX085931
   43   136.8    16.8    873  12  BI599941    BI599941 603251005
   44   136.8    16.8   4319  11  AK078681    AK078681 Mus muscu
   45   135.8    16.7
```

ALIGNMENTS

```
RESULT 1
BC025746
LOCUS       BC025746              1934 bp    mRNA    linear   HTC 04-MAR-2003
DEFINITION  Homo sapiens, clone IMAGE:5200142, mRNA.
ACCESSION   BC025746
VERSION     BC025746.1  GI:22213081
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1934)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
```

Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov

Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 49 Row: b Column: 3
This clone has the following problem: retained intron.

FEATURES   Location/Qualifiers
    source   1..1934
           /organism="Homo sapiens"
           /mol_type="mRNA"
           /db_xref="taxon:9606"
           /clone="IMAGE:5200142"
           /tissue_type="Brain, adult, 6 pooled whole brains"
           /clone_lib="NIH_MGC_114"
           /lab_host="DH10B"
           /note="Vector: pCMV-SPORT6"

ORIGIN

Query Match   100.0%;   Score 812;   DB 11;   Length 1934;
Best Local Similarity   100.0%;   Pred. No. 4.7e-222;
Matches 812;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy      1 CACTGGAGTGGGGATGGGTCCATCGGCCAACTATAAACTGATTCTCATCAGGAAACTGCACA    60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9 CACTGGAGTGGGGATGGGTCCATCGGCCAACTATAAACTGATTCTCATCAGGAAACTGCACA   68

Qy     61 TTAATCTCCCCATCACTTCAAAGGTCTTCGTCGTGGGAAGGTGACCCCAGAGGATGATTTAA   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69 TTATCTCCCCATCACTTCAAAGGTCTTCGTCAGGGGACAGAGGTGACGGCCAGGAATGATTAA   128

Qy    121 AGGTGAAAATGACAAGGTTTCCACCCCTCAAACCTTGGCTCCTTTTCTGACAAATACAGTC   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    129 AGGTGAAAATGACAAGGTTTCCACCCCTCAAACCTTGGCTCCTTTTCTGACAAATACAGTC   188

Qy    181 TGAATGAACCCGATGTCTTTTTTTTTTTACTGTGGAAATAGAGATCGGAAGAGAGTAACATTT   240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    189 TGAATGAACCCGATGTCTTTTTTTTTACTGTGGAAATAGGATCGGAAGAGAGTAACATTT   248

Qy    241 TTTTTTTTAATCCTGATAAAGAGAGATTGTTGGGAAGCTCTTTGAAAAAAAAATTTTAAATTG   300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    249 TTTTTTTAATCCTGATAAAGAGAGATTGTTGATCTTTTCCAATGAACTAATAGAGTAGTACTCT   308

Qy    301 TGGCACAGATGGATTTTTTAAAAAGTGTTGATCTTTTCCAATGAACTAATAGAGTAGTACTCT   360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    309 TGGCACAGATGGATTTTTAAAAAGTGTTAGAATCTTTCCAATGAACACTAATAGAGTACTCT   368

Qy    361 GCTCTTTGGGCTCGGATTTTTCAGGAATCGGCAATGGTGTCTCTGCAATGTCCTGGGTCCTGTAT   420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    369 GCTCTTTGGGCTCGGATTTTTCAGGAATCGGCAATGGTGTCTCTGCAATGTCCTGGGTCCTGTAT   428

Qy    421 TTGTGGATAAGTTGCTTGTGTGCAAATGCTACTCCTGCCATGGATCCCTTCAGCACCACTTCCAG   480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    429 TTGTGGATAAGTTGCTTGTGTGCAAATGCTACTCCTGCCATGGATCCCTTCAGCACCACTTCCAG   488

Qy    481 CAGCATCACCTGCACACAGAACACAGAAGAGAGGGACGTGTGTGAAGTGATAGCAGACACACCGGATGT   540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    489 CAGCATCACCTGCACAGACCAGGAGGAGACCGTGTGAAGTGATAGCAGACACACCGGATGT   548

Qy    541 TGTAACAAGAATCGGCAATTGAGAGGCGCGTCACAAACACTAAAAGTGTTCCTGCTACCTGGA   600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    549 TGTAACAAGAATCGGCATTGAGAGGCGGTCACAAACACTAAAAGTGTTCCTGCTACCTGGA   608

Qy    601 AAAGTGGCTGGGAACAACAAGAAAACGGCGCCTTCTTGCGTCGTGGATGCCTCATAGTGATTGGG   660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    609 AAAGTGGCTGGGAACAACAAGAAAACGGCGCCTTCTTGCGTCGTGGATGCCTCATAGTGATTGGG   668

Qy    661 AAATGGTGGTGTGAGGATGGGAGCCCTTGCCTAGAAGGAGAAGGAGAATGTAAAGACACTTGTTGAC   720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    669 AAATGGTGGTGTGAGGATGGGAGCCCTTGCCTAGAAGGAGAAGGAGAATGTAAAGACACTCCTGAC   728

Qy    721 AATTCTGGATGGATGTCGCAACACAGGCCAACAAAATTAAGACCACGGAGAATTCACCCAAGA   780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    729 AATTCTGGATGGATGTCGCAACACAGGCCAACAAAATTAAGACCACGGAGAATTCACCCAAGA   788

Qy    781 ACCTAACAAGAAGCATTTGTTGTGGTAGTAAAGGAA   812
          ||||||||||||||||||||||||||||||||
Db    789 ACCTAACAAGAAGCATTTGTTGTGGTAGTAAAGGAA   820
```

RESULT 2

LOCUS   AL534074   929 bp   mRNA   linear   EST 12-MAY-2003
DEFINITION   AL534074 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
           CS0DF003YG18 5-PRIME, mRNA sequence.
ACCESSION   AL534074
VERSION   AL534074.2   GI:30533521
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 929)
AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE   Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   On Feb 13, 2001 this sequence version replaced gi:12797567.
           Contact: Genoscope
           Genoscope - Centre National de Sequencage
           BP 191 91006 EVRY cedex - France
           Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
           Library was constructed by Life Technologies, a division of
           Invitrogen. This sequence belongs to sequence cluster 2511.f
           Contact :Feng Liang Email : fliang@lifetech.com URL :
           http://fulllength.invitrogen.com/. InVitroGen Corporation 1600
           Faraday Avenue Genoscope sequence ID : CS0DF003BD09QP1.

FEATURES   Location/Qualifiers
    source   1..929
           /organism="Homo sapiens"
           /mol_type="mRNA"
           /db_xref="taxon:9606"
           /clone="CS0DF003YG18"
           /tissue_type="FETAL BRAIN"
           /dev_stage="fetal"
           /clone_lib="Homo sapiens FETAL BRAIN"
           /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
           was primed with a Not1-oligo(dT) primer. Five prime end
           enriched, double-strand cDNA was digested with Not I and
           cloned into the Not I and EcoRV sites of the pCMVSPORT 6
           vector. Library was not normalized."

ORIGIN

Query Match   98.5%;   Score 800.2;   DB 9;   Length 929;
Best Local Similarity   99.6%;   Pred. No. 1e-218;
Matches 810;   Conservative   2;   Mismatches   0;   Indels   1;   Gaps   1;

```
Qy      1 CACTGGAGTGGGGATGGGTCCATCGGCCAACTATAAACTGATTCTCATCAGGAAACTGCACA    60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CACTGGAGTGGGGATGGGTCCATCGGCCAACTATAAACTGATTCTCATCAGGAAACTGCACA   60

Qy     61 TTAATCTCCCCATCACTTCAAAGGTCTTCGTCGTGGGAAGGTGACCCCAGAGGATGATTTAA   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TTATCTCCCCATCACTTCAAAGGTCTTCGTCAGGGGACAGAGGTGACGGCCAGGAATGATTTAA   120

Qy    121 AGGTGAAAATGACAAGGTTTCCACCCCTCAAACCTTGGCTCCTTTTCTGACAAATACAGTC   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGGTGAAAATGACAAGGTTTCCACCCCTCAAACCTTGGCTCCTTTTCTGACAAATACAGTC   180

Qy    181 AGGTGAAAATGACAAGGTTTCCACCCCTCAAACCTTGGCTCCTTTTCTGACAAATACAGTC   245
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    186 AGGTGAAAATGACAAGGTTTCACCCCCTCAAACCTTGGCTCCTTTTCTGACAAATACAGTC   245
```

```
Qy    181  TGAATGAACCGGATCTCTTTTTTTTTTTTTTTACTGTGGAAATAGGATAGGATTAGGATCATTT  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    246  TGAATGAACCGGATGTCTTTTTTTTTTTTTTTACTGTGGAAATAGGATCGGAAATAGGATAACATTT  305

Qy    241  TTTTTTTAATCCTGATAAAGAAGAGATTGTTGGGAAGCTCTTTGAAAAAGAAATTTAAAATTG  300
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    306  TTTTTTTAATCCTGATAAAGAAGAGATTGTTGGGAAGCTCTTTGAAAAAGAAATTTAAAATTG  365

Qy    301  TGGGACAGATGGATTTAAAAAGTGTTAGATCTTTCCAATGAACACTAATAGAGTACTCT  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    366  TGGGACAGATGGATTTAAAAAGTGTTAGATCTTTCCAATGAACACTAATAGAGTACTCT  425

Qy    361  GCTCTTGGCTGATTTTTCAGAGAATGCAATGGTTCTCTGCGATGTCCTTGGGTGTTAT  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    426  GCTCTTGGCTGATTTTTCAGAGAATGGCAATGGTTCTCTGCGATGTCCTTGGGTCCTGTAT  485

Qy    421  TTGTGGATAAGTGCTTGTGCAATGCTACTCTGCCATGGATCCCTTCAGCACACTTTCCAG  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    486  TTGTGGATAAGTGCTTGTGCAATGCTACTCTGCCATGGATCCCTTCAGCACACTTTCCAG  545

Qy    481  CAGCATCACCTGCACAGAAGGTCTCGTCAGGAGGGTGACGGCCAGAGATGATTTAAAG  540
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    546  CAGCATCACCTGCACAGAAGGTCTCGTCAGGAGGGTGACGGCCAGAGATGATTTAAAG  605

Qy    541  TGTAACAAGAATCGCATTGAGGGAGGG-GTCACAAACAGTAAAGTGTTCACCTTCTACCTGG  599
           ||||||||||||||||||||||||||||||||||||||||||||
Db    606  TGTAACAAGAATCGCATTGAGGGAGGCGGTCACAAACAGTAAAGTGTTCACTABAGTCTG  665

Qy    600  AAAAGTGGCTGGAACAACAAGAAAACCGGCCTTCGTGCGTCAGTAGTTGGAAGTAACA  659
           ||||||||||||||||||||||||||||||||||||||||||||
Db    666  AAAAGTGGCTGGAACAACAAGAAAACCGGCCTTCGTGCGTCAGTAGTTGGAAGTAACA  725

Qy    660  GAAATGGTGGTTGTGGATGAGCCTTGCCTTAGAAGGAAGAGGAAGTAATCTTAGAGTACT  719
           ||||||||||||||||||||||||||||||||||||||||||||
Db    726  GAAATGGTGGTTGTGGATGAGCCTTGCCTTAGAAGGAAGAGGAAGTAATCTTAGAGTACT  785

Qy    720  CAATTCTGGATGGATGGGCCAACAGGCACACAAAATTAAGACCACGAGAATTCACCCAAG  779
           ||||||||||||||||||||||||||||||||||||||||||||
Db    786  CAATTCTGGATGGATGGGCCAACAGGCACACAAAATTAAGACCACGAGAATTCACCCAAG  845

Qy    780  AACCTAACAGAAGCAATTTGTGGTAGTTAAAGGAA  812
           |||||||||||||||||||:||||:|||||||||
Db    846  AACCTAACAGAAGCAATTTGTGGKAGTAAAGGAA  878
```

```
ORIGIN
```

```
RESULT 3
LOCUS       BX460561       940 bp    mRNA      linear    EST 22-MAY-2003
DEFINITION  BX460561 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF015YL24 5-PRIME, mRNA sequence.
ACCESSION   BX460561
VERSION     BX460561.1  GI:31035211
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 940)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2511.f,
            Contact: Feng Liang Email: fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF015DF12Qt1.
FEATURES             Location/Qualifiers
     source          1..940
                     /organism="Homo sapiens"
```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF015YL24"
/tissue_type="FETAL BRAIN"
/clone_lib="Homo sapiens FETAL BRAIN"
/dev_stage="fetal"
/note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
was primed with a NotI-oligo(dt) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

```
ORIGIN
```

Query Match   94.5%; Score 767.2; DB 13; Length 940;
Best Local Similarity   97.2%; Pred. No. 3.1e-209;
Matches 791; Conservative 10; Mismatches 9; Indels 4; Gaps 2;

```
Qy      3  CTGGAGTGGGGATGGGTCCATGCGGCAACTATAAACTGATTCTCATCAGGAAAAACTGCACATT  62
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     50  CCGGGAAATTCCCGGGGATCCATGCGGCAACTATAAACTGATTCTCATCAGGAAAAACTGCACATT  109

Qy     63  ATCTCCCCATCACTTCAAAAAGGTCTCGTCAGGCAGCAGAGGGTGACGGCCAGAGATGATTTAAAG  122
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    110  ATCTCCCATCACTTCAAAAAGGTCTCGTCAGGCAGCAGAGGGTGACGGCCAGAGATGATTTAAAG  169

Qy    123  GTGAAAATGACAAGGTTTCCACCCCTCAAAACCTTGGCGTCCTTTTCTGACAATACAGTCTG  182
           |||||||||||||||||||||||||||||||||||||||||||||||
Db    170  GTGAAAATGACAAGGTTTCCACCCCTCAAAACCTTGGCGTCCTTTTCTGACAATABAGTCTG  229

Qy    183  AATGAACCCGGATGTCTTTTTTTTTTTTACTGTGGAAATAGGATCGGAAAGTAACA----TT  239
           ||||||||||||||||||||||||||||||||||||||||||    ||
Db    230  AATGAACCCGGATGTCTTTTTTTTTTTTACTGTGGAAATAGGATCGGCAATABAGTCTG-TT  289

Qy    240  TTTTTTTAATCCTGATAAAGA-AGATTGTTGGGAAGCTCTTTGAAAAAGAAATTTTAAAT  298
           |||||||||||||||| |||||||||||||||||||||||||||||||
Db    290  TTTTTTTAATCCTGATAAAGATAGATTGTTGGGAAGCTCTTTGAAAAAGAAATTTTAAAT  349

Qy    299  TGTGGCACAGATGGATTTAAAAAGTGTTAGATCTTTCCAATGAACACTAATAGAGTACT  358
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    350  TGTGGCACAGATGGATTTAAAAAGTGTTAGATCTTTCCAATGAACACTAATAGAGTACT  409

Qy    359  CTGCTCTTGGCTGATTTTTCAGAGAATGGCAATGGTTCTCTGCGGGTCCTGT  418
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    410  CTGCTCTTGGCTGATTTTTCAGAGAATGGCAATGGTCTCTGCCATGTCCCTGT  469

Qy    419  ATTTGTGGATAAGTGCTTGTGCAATGCTACTCTGCCATGGATCCCTTCAGCACACTTCC  478
           |||||||||||||||||||||||||||||||||||||||||||||||
Db    470  ATTTGTGGATAAGTGCTTGTGCAATGCTACTCTGCCATGGATCCCTTCAGCACACTTTCC  529

Qy    479  AGCAGCATCACCTGCACAGACCAGAAGAGGAGGACGTGTGAAGTAGCAGCAACACCGAT  538
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    530  AGCAGCATCACCTGCACAGACCAGAAGAGGAGGACGTGTGAAGTAGCAGCAACACCGAT  589

Qy    539  GTTGTAACAAGAATCGCATTGAGGAGCGCGTCACAAACAGTAAAGTGTTCCTGTCTTACCTG  598
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    590  GTTGTAACAAGAATCGCATTGAGGAGCGCGTCACAAACAGTAAAGTGTTCCTGTCTTACCTG  649

Qy    599  GAAAAGTGGCTGGAACAACAAGAAAACCGGCCTTCGTGCGTCAGTAGTTCCATAGTGATTG  658
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    650  GAAAAGTGGCTGGAACAACAAGAAAACCGGCCTTCGTGCGTCAGTAGTTCCATAGTGATTG  709

Qy    659  GGAAATGGTGGTTGTGTGGATGGAGCCTTGCCTTAGAAGGAAGAGGAAGTAAGACACTCCCTG  718
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    710  GGAAATGGTGGTTGTGTGGATGGAGCCTTGCCTTAGAAGGAAGAGGAAGTAAGACACTCCCTG  769

Qy    719  ACAATTCTGGATGGATGTGGGCCWAMAGGCAAMAAAATTAAGAMCAMGAGAATTCACCCAA  778
           |||||||||||:||||||||||||||||||||||||||||||||
Db    770  ACAATTCTGGATGGAGTTGGGCCWAMAGGCAAMAAAATTAAGAMCAMGAGAATTCACCCAA  829

Qy    779  GAACCTAACAGAAGCAATTTGTGGTAGTAAAGGAA  812
           ||||||||||||||||||:||||:|||||||||
Db    830  GAACCTAACAGAAGCAATTTKTGGTAGTAAAGGAA  863
```

RESULT 4

LOCUS   BI756155   793 bp   mRNA   linear   EST 25-SEP-2001
DEFINITION 603030509F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5200142 5',
            mRNA sequence.
ACCESSION   BI756155
VERSION   BI756155.1 GI:15747733
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 793)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11501 row: f column: 15
            High quality sequence stop: 793.
FEATURES   Location/Qualifiers
     source   1..793
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5200142"
                     /clone_lib="NIH_MGC_114"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon cloning). Average insert size 1.5 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was constructed by C. Gruber
                     (Invitrogen). Research Genetics tracking code 019. Note:
                     this is a NIH_MGC Library."
ORIGIN

Query Match   82.1%; Score 667; DB 12; Length 793;
Best Local Similarity 98.2%; Pred. No. 1.9e-180;
Matches 769; Conservative 0; Mismatches 5; Indels 9; Gaps 9;

Qy   1 CACTCGAGTGGGGATGGTCCATCGGGCAACTATAAACTGATTCTCATCAGGA-AACTGCAC 59
           ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   9 CACTCGAGTGGGGATGGTCCATCGGGCAACTATAAACTGATTCTCATCAGGAAAACTGCAC 68

Qy   60 ATTATCTCCCCATCACTCACTCAAAAGGTCTCGTCAGCAGAGGTGACGCCAGGAGATGATTTA 119
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69 ATTATCTCCCCATCACTCACTCAAAAGGTCTCGTCAGCAGAGGTGACGCCAGGAGATGATTTA 128

Qy   120 AAGGTGAAAAATGACAAGGTTTCCACCCCCTCAAACACTTGGCTCCTTTTTCTGACAATACAGT 179
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   129 AAGGTGAAAAATGACAAGGTTTCCACCCCCTCAAACACTTGGCTCCTTTTTCTGACAATACAGT 188

Qy   180 CTGAATGAAACCCGATGTCTCTTTTTTTACTGTGGAAAATAGGATCGGAAAGAGAGTAACATT 239
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   189 CTGAATGAAACCCGATGTCTTTTTTTTTTTTTACTGTGGAAAATAGGATCGGAAAGAGAGTAACATT 248

Qy   240 TTTTTTTTAATCCTGATAAAGAAGAGATTGTTGTTGGGAAGCTCTTTTGAA-AAAATTTTAAT 298
           ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   249 TTTTTTTTAATCCTGATAAAGAAGAGATTGTTGTTGGGAAGCTCTTTTGAAAAAAAATTTTAAT 308

Qy   299 TGTGGCACAGAGATGGGATTTTAAAAAGTGTTAGATCTTTCCAATGAACACTAATAGAGTACT 358
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   309 GGTGGCACAGATGGGATTTTAAAAAAGTGTTAGATCTTTCCAATGAACACTAATAGAGTACT 368

Qy   359 CTGCTCTTCTTGGCTTGGATTTTTCAGAGAATGGGCAA-TGGTCTCTGCCGATGTCCTGTGGGTCCTG 417
           |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   369 CTGCTCTTCTTGGCTTGGATTTTTCAGAGAATGGGCAAGTGGTCTCTGCCGATGTCCTGTGGGTCCTG 428

Qy   418 TATT-TGTGGATAAGTGCTTGTGCAATGCTACTCTGCCATGGATCCCTTCAGCACACTT 476
           |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   429 TATTTTGGGGGATAAGTGCTTGTGCAATGCTACTCTGCCATGGATCCCTTCAGCACACTT 488

Qy   477 CCAGCACCATCACCTCGACAGACCAGAGAAGAGGAGGAGGTGTGAAAGTGAATAGCAGACACCG 536
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   489 CCAGCACGCATCACCTCGACAGAGCCAGAGAAGAGGAGGAGGTGTGAAAGTGAATAGCAGACACCG 548

Qy   537 ATGTTGTTAACAAGAATGCATTGAGAGCGGTCACAAACAGTAAA-GTGTTCCTGTTGTCTAC 595
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   549 ATGTTGTTAACAAGAATCGCATTGAGAGCGGTCACAAACAGTAAAGTGTTCCTGTTGTCTAC 608

Qy   596 C-TGGAAAAGTGGCTGGAACAACAAGAAAACCGGCCTTCTTGGCTTCCATAGTG 654
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   609 CGTGGAAAAGTGGCTGGAACAACAAGAAAACCGGCCTTCTTGGCTTCCATAGTG 668

Qy   655 ATTGGGAAAATGGTGGTGTGAGATGGGACCCTTGCCTAGAAGAGAGAAATGTAAGAGACTC 714
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   669 ATTGGGAAAATGGTGGTGTGAGATGGGACCCTTGCCTAGAAGAGAGAAATGTAAGAGACTC 727

Qy   715 CCTGACAATTCTGGATGGTGCGCAACCAACAAAAATTAAGACCACCAGAGAATTCAC 774
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   728 CCTGACAA-TCTGGATGGATGTGCGCAACCAACAAAAATTAAGACCA-GAGAATTCAC 785

Qy   775 CCA 777
           |||
Db   786 CAA 788

RESULT 5

LOCUS   BI488689   1022 bp   mRNA   linear   EST 28-AUG-2001
DEFINITION 603021235F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5191688 5',
            mRNA sequence.
ACCESSION   BI488689
VERSION   BI488689.1 GI:15327917
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1022)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11479 row: f column: 09
            High quality sequence stop: 680.
FEATURES   Location/Qualifiers
     source   1..1022
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5191688"
                     /clone_lib="NIH_MGC_114"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT

```
RESULT 6

Query Match    75.1%; Score 610.2; DB 12; Length 1022;
Best Local Similarity   97.6%;
Matches  584; Conservative  0; Mismatches  8; Indels  9; Gaps  6;

Qy    1 CACTGGAGTGGGGATGGTCCATCGGCCAACTATATAAACTGATTCTCATCAGGAAACTGCACA   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   12 CACTGGAGTGGGGATGGTCCATCGGCCAACTATATAAACTGATTCTCATCAGGAAACTGCACA   71

Qy   61 TTACTCCCCATCACTTCAAAGGTCTCGTCAGCGAGAGGTGACCCCAGAGATGATTTAA      120
        |||||||||||||||||||||||  |||||||||||||||||||||||||||||||||||
Db   72 TTACTCCCCATCACTTCAAAGGTTCGTCAGCGAGAGGTGACCCCAGAGATGATTTAA       131

Qy  121 AGGTGAAAATGACAAGGTTTCCACCCCTCAAAACCTTGGCTCCTTTTCTGACAATACAGTC    180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  132 AGGTGAAAATGACAAGGTTTCCACCCCTCAAAACCTTGGCTCCTTTTCTGACAATACAGTC    191

Qy  181 TGAATGAACCCGATGTCTCTTTTTTTTTACTGTGGAAATAGGATCGGAAGAGAGTAACATTT   240
        |||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db  192 TGAATGAACCCGATGTCTTTTTTTTTTACTGTGGAAATAGGATCGGAAGAGAGTAACATTT    251

Qy  241 TTTTTTTTTTAATCCTGATAAAGAGAAATTGTTGGGAAGCTCTTTGAAAAAAATTTTAAATTG  300
        |||||||||   |||||||||||||||||||||||||||||||||||||||||||||||||
Db  252 TTTTTTTTTAA CCTGATAAAGAGAAATTGTTGGGAAGCTCTTTGAAAAAAATTTTAAATTG   311

Qy  301 TGGACAGATGGATTTAAAAAGTGTTAGATCTTTCCAATGAACACTAATAGAGTACTCT      360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  312 TGGACAGATGGATTTAAAAAGTGTTAGATCTTTCCAATGAACACTAATAGAGTACTCT     371

Qy  361 GCTCTCTTGGCTGGATTTTTTCAGAGAATGGCCAATGGTTCTCTCGCGGATGTCCTGTGTAT   420
        |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db  372 GCTCTCTTGGCTGGATTTTTTCAGAGAATGG CCAATGGTTCTCTCGCGGATGTCCTGTGTAT  431

Qy  421 TTGTGGGATAAAGTGCTTGTTGCAA TGCTACTCTGGCCATGGAATCCCTCAGCACACATTTCCA  479
        |||||||||||||||||||||||||  |||||||||||||||||||||||||||||||||||
Db  432 TTGTGGGATAAAGTGCTTGTTGCAATTGCTACTCTGGCCATGGAATCCCTCAGCACACTTTCCA  491

Qy  480 GCAGCATCACCTGACAGACCAGAGGAGGGGACCTGTGAAGTGATAGCAGCACACCGATG      539
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  492 GCAGCATCACCTGACAGACCAGAGGAGGGGACCTGTGAAGTGATAGCAGCACACCGATG     548

Qy  540 TTGTAACAAGAAATCCATTGAAGGAGCGGGTCA CAAACAGTAAAGTAGTTCCTGTTGTTACCTG  598
        |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db  549 TTGTAACAAGAAATCCATTGAAGGAGCGGGTCAACAAACAGTAAAGTAGTTCCTGTTGTTACCTG  608

Qy  599 GAAAAGTGGGCTGGAA  CAACAAGAAAACCGGCGCCTTCTTCGGTCGATGCCCATAGTGAT     656
        ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db  609 GAAAAGTGGGCTGGAAGACAACAAGAAAACCGGCGCCTTCTTCGGTCGATGCCCATAGTGAT    668

Qy  657 TGGGAA ATGGTGCTGTG AGATGGGAGCCCTTGCCTTAGAAGG               695
        |||||| ||||||||||| ||||||||||||||||||||||||
Db  669 TGGGAACATGGTGCTGTGTGCAGATGGGAGCCCTTGCCTTAACAGG            709
```

```

Note at top of GenBank header:

> primed and directionally cloned (EcoRV site is destroyed
> upon cloning).  Average insert size 1.5 kb, insert size
> range 1-3 kb.  Library is normalized and enriched for
> full-length clones and was constructed by C. Gruber
> (Invitrogen).  Research Genetics tracking code 019. Note:
> this is a NIH_MGC Library."

```
ORIGIN

          /db_xref="FANTOM_DB:C630045F22"
          /db_xref="MGI:2477712"
          /db_xref="taxon:10090"
          /clone="C630045F22"
          /sex="male"
          /tissue_type="hippocampus"
          /clone_lib="RIKEN full-length enriched mouse cDNA library"
          /dev_stage="adult"
          1..3258
          /note="unclassifiable"

misc_feature

Query Match   67.9%; Score 551.6; Pred. No. 4e-147; DB 11; Length 3258;
Best Local Similarity   82.5%; Conservative   0; Mismatches 109; Indels   38; Gaps   4;
Matches 695;

Qy   2 ACTGGAGTGGGAGGTCCATCGCGCAACTATAAAACTGATTCTCATCAGAGAAACTGCACAT 61
          ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   31 ACTGGAGTAGGGATGGTCTGTCCGTGGCTATAAAACTGCTTCATCAGAGAATGCAGAAACT--ACAT 88

Qy   62 TATCTCCCCATCACTTCAAAAGGTCTCGTCGTCAGGCAGAGAGGTGACCCCAGGAGATGATGAGTTTAAA 121
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   89 TAGCTCACCATCACTTCAAAAGGTCTCGTCGTCAGACAGAGAGGTGACCCCAGGGGATGATTTTAA 148

Qy   122 GGTGAAAATGACAAGGTTTCCACCCCTCAAACCTTGGCTCCTTTTCTGACAATACAGTCT 181
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   149 GGTGAACATGACAAGGTTTCACCCCTCAAAACCTTGCTTCCTTTTCTGACAATACAG--T 206

Qy   182 GAATGAACCCGATGTCTTTTTTTTACTGTGGAAATAGGATCGGAAAGAGAGTAAC----- 236
          |||||||||||| |||||||||||||||||||||||||||||||||||
Db   207 GAACGAGCCCAACGTC---TTCACACTGTGGAAACGGGACTGAAAGAGAAAAAGCTTGCC 263

Qy   237 ------------------ATTTTTTTTTAATCCTGATAAAAGAAGATTGTTG 270
                          ||||||||||||||||||||||||||||||
Db   264 TTTTTTTCTTTTTTCTTTTCTTTTTCAATTCTTAATAAAAGAATTGCTGGG 323

Qy   271 GGAAGCTCTTTGAAAAAAAATTTTAAATGTGGGACAGAGTGGATTTAAAAAGTGTTAGA 330
          ||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   324 AAGCATTCTCTTTGAAAAATCTCAGAACTGGCACAGATGGATTTATAAAAGTGTTAGC 383

Qy   331 TCTTTCCAATGAACAACTAATAGAGTACTCGCTCTTGGCTGGAATTTTTTCAGAGAATGGCA 390
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   384 TCTTTCCAATGAGACACTAGAGGGTTCCCTCGCTCTTGGCTGGAATTTTTCAGGAATGGCA 443

Qy   391 ATGGTCTCTGGGATGTCCTGGGTCTGTATTGTGTGGATAAGTGGTGTGCAATGCTACTC 450
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   444 ATGGTCTCTGCAAGTCCTGGGCCCTGTACTTGTGGAATAAGTGCTTTGTGCCGGATCCTGCTC 503

Qy   451 TGCCATGGATCCCTTCAGCACACACTTTCCAGCAGATCACCCTGCACAGACCAGAGAGGG 510
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   504 TGCCACGGGGTCACTCCAACACACACTTCAGCGCAGATCACCTGCACGGTCACTCCAGAGGAGG 563

Qy   511 ACGTGTGAAGTGATAGCAGCAGCACACGGATTGGTAACAAGAAATCGCATTGAGGACGGGTCA 570
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   564 ACCTGTGAAGTGAATCGGCGGCCCACAGTGTTGTAACAAGAAACCCATCCAGGACGGGTCA 623

Qy   571 CAAACAGTAAAGTGTTCCTGTCTACCTGGAAAAGTGGCTGGAACAACACACAAACCGGACCT 630
          ||||||||||||||||||||||||||||||||||||||||||||
Db   624 CAAACAGTGAAGTGTTCCTGTTACCTGGGAAAGTGGCTGGAACAACAGAAAACCGACCT 683

Qy   631 TCTTGCCGTGTCGATGCCTCCATAGTGAGTTGGTGGAAATGGTTGGTGTGAGATGAGAGCCTTGCCTA 690
          ||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   684 TCCTGTGTCGGATGCCTCCATAGTAATTGGGAAAATGGTGGAGAGGCCCCTGCCTA 743

Qy   691 GAAGGAGAGAAGAATGTAAGACACTCCCTGACACAATTCTTGGATGGATGGGCCAACAGGCCAAC 750
          ||||||||||||||||||||||||||||||||||||||||||||
Db   744 GAAGGAGAGAATGTAAGACTCCCTGACAATTCTGGATGGATGTGCGGCAACAGGGCAAC 803

Qy   751 AAAATTAAGACCACGAGAAATTCACCCAAGAGAACCTAACAGAAGAACCTAACAGAGAAGG 810
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   804 AAGAATTAAGACTACAGAAATTCACCCAAGAGAACCTAACAGAAGAAGCATTGTTGG 863

Qy   811 AA 812
          ||
Db   812
```

```
                                                 |
Db    864 GA  865
```

```
RESULT 7
LOCUS       CK001752                  672 bp    mRNA    linear    EST  26-NOV-2003
DEFINITION  AGENCOURT_16385450 NIH_MGC_227 Homo sapiens cDNA clone
            IMAGE:30718008 5', mRNA sequence.
ACCESSION   CK001752
VERSION     CK001752.1  GI:38527786
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 672)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Professor Miklos Palkovits
            cDNA Library Preparation: Michael Brownstein / Ted Usdin
            Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM260  row: j  column: 01
            High quality sequence stop: 579.
FEATURES             Location/Qualifiers
     source          1..672
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30718008"
                     /tissue_type="Bulk tissue from Human Spinal cord"
                     /clone_lib="NIH_MGC_227"
                     /lab_host="DH10B TonA"
                     /note="Organ: brain/CNS; Vector: pDNR-LIB; Site 1: SfiI
                     (directional); Site 2: SfiI (directional); Library is
                     oligo-dT primed and directionally cloned. 5' and 3'
                     adaptors were used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3'
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-d(T)30N-1N-3. Full-length
                     enriched library was constructed using the Clontech
                     Creator SMART kit and size-selected for >0.5kb with an
                     average insert size of 1.3kb library created in the
                     laboratory of Jonathan Kuo and Ted Usdin."

ORIGIN

Query Match    62.9%;  Score 511;  Pred. No. 1.4e-135;  DB 14;  Length 672;
Best Local Similarity   100.0%;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
Matches  511;

Qy     91 AGGCAGAGGTGACGCCAGGAGATGATTTTAAAGGTGAAAATGACAAGGTTTCCACCCCTCA  150
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4 AGGCAGAGGTGACGCCAGGAGATGATTTTAAAGGTGAAAATGACAAGGTTTCCACCCCTCA  63

Qy    151 AACCTTGGCTCCTTTTCTGACAATACAGTCTGAATGAATGAACCCGATGTCTTTTTTTACTG  210
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db     64 AACCTTGGCTCCTTTTCTGACAATACAGTCTGAATGAATGAACCCGATGTCTTTTTTTACTG  123

Qy    211 TGGAAATAGGATCGGAAAGAGAGTAACATTTTTTTTTTTTTTAATCCTGATAAAAGAAGTTGTTG  270
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    124 TGGAAATAGGATCGGAAAGAGAGTAACATTTTTTTTTTTTTAATCCTGATAAAAGAAGATTGTTG  183

Qy    271 GGAAGCTCTTTGAAAAAAAATTTTAAATGTGGCACAGATGGATTTAAAAAGTGTGTTAGA  330
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db         330
```

RESULT 8

LOCUS   BU118807   768 bp   mRNA   linear   EST 25-NOV-2002
DEFINITION 603143059F1 CSEQCHL16 Gallus gallus cDNA clone ChEST137f15 5', mRNA sequence.
ACCESSION   BU118807
VERSION   BU118807.1 GI:25327474
KEYWORDS   EST.
SOURCE   Gallus gallus (chicken)
  ORGANISM Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1 (bases 1 to 768)
  AUTHORS   Boardman,P.E., Sanz-Ezguerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
  TITLE   A Comprehensive Collection of Chicken cDNAs
  JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
  PUBMED   12445392
  MEDLINE   22335534
COMMENT   Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology
            (UMIST)
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
FEATURES   Location/Qualifiers
     source   1..768
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="Compton Line 151"
                     /db_xref="taxon:9031"
                     /clone="ChEST137f15"
                     /sex="Female"
                     /tissue_type="not cerebrum or cerebellum"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHL16"
                     /note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site_2: NotI; Modification of pBluescript II KS(+)
                     [Stratagene] vector to accommodate cDNA produced with the
                     T-trimmed protocol (Construction of uni-directionally
                     cloned cDNA libraries from messenger RNA for improved 3'
                     end DNA sequencing by Glenn Fu, et al. U.S. Patent #
                     6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI.
                     Ligate in double stranded adaptor containing BsgI and
                     BamHI sites [5'ggccgcgtgcgagcccggatccgaaaaaaag]
ORIGIN   [5'aattctttttcggatccgggctgcacgc]"

Query Match   58.5%; Score 475.4; DB 13; Length 768;
Best Local Similarity   82.3%; Pred. No. 2.4e-125;
Matches 572; Conservative 0; Mismatches 116; Indels 7; Gaps 2;

Qy   98 GGTGACGCCAGGAGGATGATTAAAGGTGAAAATGACAAGGTTCACCCTCAAACCTTG 157
          |||||| || | || ||| ||| |||||||| |||| || |||| |||||||||||
Db   10 GGAGGCGGGGCGCCGACGAGGGTTAAGCTGAAAAATGAGAAAGGTTCCACCCTCAAACCTCG 69

Qy   158 GCTCCTTTCTCAGCACACATTTCAGCAGTCTGAATG--AACCCGATGTCTTTTTT 215
          || ||||| |||| ||| ||||||||||||| || | ||||| ||||| ||||
Db   70 GCCACTTTTCTGACGGGCAGTGCCAGTGCGTCTGAGGTCGTGTCTTTACTGTGTGGAA 129

Qy   216 ATAGGATCGGAAGGAGTAACATTTTTTTTTTGTCCTCTGTAATCCTGAATAAAGGAG 275
          ||| |||||| ||| || |||| |||||||| |||| |||| | ||||| ||| |||
Db   130 TTTGTAACAAGAGATTTTTTTTTTTGTCCTCTGTAATCCTGAATTAAAGGAAGAA 189

Qy   276 CTCTTTGAAAAAAATTTTAAAATTGTGGCACAGATGGGATTTTTAAAAGTGTTTAGATCTTT 335
          ||||| |||| |||||||| |||||| || ||||||| |||||||||||| ||||
Db   190 CTATT-----AAAAATCCCTTAITGIGGTGCAAGATGGATTTTTAAAAAGTGTTAGACTTTT 244

Qy   336 CCAATGAACACTAATAGAGTACTCCTGCTCTTGGCTGGATTTTTCAGAGAATGGCAATGGT 395
          |||||| ||| |||| ||||| ||||||||| |||||||||| ||| ||||| |||||||
Db   245 CCAATGAACACTAATAGAGTGCTTCCTGGCTCGATTACTCAGAGAATGGCAGTGGT 304

Qy   396 CTCTGCGATGTCCTGGGGTCCTGTAITTGTGTGGATAAGTAAGTGTCAATGCTACTCTGCCA 455
          |||| ||| ||||| ||| || || || |||| |||||||||||| ||||| |||||||
Db   305 CTCTGCAATGTCCTGGGTCCGTATTTGTGTGGATAAGTGCAATGCTCGTCTCGTCA 364

Qy   456 TGGATCCTTCAGCACTTTCAGCAGACCATCACCTGCACAGATCCCAGAAGGAGGGACGTG 515
          ||||| |||| ||| ||||| |||||| ||||||| ||||| || ||||||| |||||
Db   365 GGGATCCTTCAGCACTTTCCGCAGACAGCCAGAAGGAGGAAACGTG 424

Qy   516 TGAAGTGATAGCAGCAGCACCGATGCTTGTAACAAGAAATCGCAITTGAGAGCGTCACAAAC 575
          |||||| |||||||||| |||||||||| || | |||||| ||||||||||||| |||||
Db   425 TGAAGTTATAGCAGCACACAGAGATGCTGTAATAAGAAATGCAAAGATCACACAAAC 484

Qy   576 AGTAAAGTGTTCCTGTCTACCTGGAAAAAGTGGCTGGAACAACAAGAAACCGGCCTTCCTG 635
          || ||| ||||||| | |||||||||||||| |||||| ||||||| || ||||||||||
Db   485 AGTAAAATGCTCCTGTGTTCTACCTGGGGAAAGTGGCTGGACTACAGACAGACCTTCCTG 544

Qy   636 CGTCGGATGCCTCCATAGTGATTTGGGAAATGGTGTGTTGTGTTGAGATGGAGG 695
          || |||||| |||| ||||| || |||||||| || || ||||| || |||||
Db   545 TGTTGATGCTTCAATAGTGATTGGGAAAATGGTGTGTGATGGAGGCCTGCCTAGACGG 604

Qy   696 AGAGGAAGTTAAGACACTCCCTGACAAATTCTGGATGAATGGTGCGCACAGGCGAACAAAAT 755
          |||||| |||||| ||||||||||| ||||||| |||||||||||| |||||||| |||||
Db   605 AGAGGAAGTAGAGACAGTTGCCGTTGGTGCAATTCTGGATGAGTGTGTGCGGACTGGCAATAAAT 664

Qy   756 TAAGACCACGAGAATTCACCCAGAGAAACCTAACAGA 790
          ||| |||||| |||||||| ||||||| ||||||||
Db   665 CAAGACCACCAGAATCCACCCAAGGACCTAACAGA 699

ORIGIN

RESULT 9

LOCUS   BF652347   506 bp   mRNA   linear   EST 25-APR-2001
DEFINITION 275877 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF652347
VERSION   BF652347.1 GI:11917477
KEYWORDS   EST.
SOURCE   Bos taurus (cow)
  ORGANISM Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 506)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
            Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
            Chitko-McKown,C.G., Pertea,G., Holt,I., Karamycheva,S., Liang,F.,

TITLE   Quackenbush,J. and Keele,J.W.
            Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
JOURNAL   Genome Res. 11 (4), 626-630 (2001)
MEDLINE   21180013
PUBMED   11282978
COMMENT   Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt trimmed with phred
            and -minscore 18. Vector identified by cross_match with the -minscore 18
            and -mismatch 12 options.
            PCR Primers
            FORWARD: AGGAAACAGCTAITGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 65 row: M column: 8
            Seq primer: ATTTAGGTGACACTATAG.

FEATURES   Location/Qualifiers
     source   1..506
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 3BOV"
                     /note="Vector: PCMV SPORT6; Site 1: NotI; Site 2: SalI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."

ORIGIN

Query Match   54.5%;   Score 442.2;   DB 10;   Length 506;
Best Local Similarity 93.3%;   Pred. No. 7.7e-116;
Matches 473;   Conservative 0;   Mismatches 33;   Indels 1;   Gaps 1;

Qy   249 ATCCCTGATAAAGAAGAGATTGTTGGGGAAGCTCTTGAAAAAAAATTTAAATTGTGGCACAG 308
             ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATCCCTGATAAAGAAGA-TATTGGGGAAGCTCCTTTTAAAAAATCTTCAATTGTGGCACAG   59

Qy   309 ATGGATTTAAAAAGTGTTAGATCTTTCCAATGAACCATAATAGAGTACTCTGCTCTTCTGG 368
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   60 ATGGATTTAGAAAGTGTTAGAGCTTTCCAATGAACACTAATAGAGTACTCTGCTCTTCTGG 119

Qy   369 CTGGATTTTTCAGAGAATGGGCAAATGGTCTCTTGCGATGTCTGGGTCCTGTATTTGTGGAT 428
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   120 CTGGATTTTTCAGAGAATGGGCAAATGGTCTCTCGCGATGTCTGGGTCCTGTATTTGTGGAT 179

Qy   429 AAGTGCTTGTTCGCAATGGTACTCTGCCATGGATCCCTTCAGCACTTTTCAGCGACCATCA 488
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   180 AAGTGCTTGGCGGTGTCTTCTTGCCATGGATCCCTTCAGCACTTTTCAGCGACCATCA 239

Qy   489 CCTGCACAGACCAGAAGGAGGGACGTGCTGAAGTGATAGAGCACACCGGATTGTGAACA 548
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   240 CCTGCACCGACCAGAAGGAGGGACGTGTGAAGTGATTGCTCGCGCACAGATGTTGTAATAA 299

Qy   549 GAATCCCATTGAGACGGGTCACAAAACAGTAAAAGTGTTCCTTGGTCTACCTGAAAAGTGGC 608
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   300 AAAACCGCATTGAGGAGGGGTCGCAAAACAGTCAAGTGTTCCGTGTCTGCTTGGGAAAGTGGC 359

Qy   609 TGGAACAACAAGAAACCGGCCTTCTTCTTGCGTCGAATCGCTCCATAGTTGGGAAAATGGGTG 668
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   360 TGGAACAACAAGAAACCGGACACTTCCTCGTGGTGATGCCTCATTGTGATCGGGGAAATGGGTG 419

Qy   669 GTGTGAGATGGAGCCCTTGCCTAGAGAGGAGGAGAGAATGTAAGACACTCCCTGACAATTCTGG 728
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   420 GTGTGAGAGTGGAGCCCTTGTTCTTAGAGGAGGAAGAGTGTAAGACACTCCCTGATAATTCTGG 479

Qy   729 ATGGAGTGTGCCAACAGCAACAAAT 755
             ||||||||||||||||||||||||
Db   480 ATGGAGTGTGCTTACAGGCAATAAAAT 506

RESULT 10
LOCUS   BU615139   706 bp   mRNA   linear   EST 20-FEB-2003
DEFINITION UI-M-FR0-cbe-1-08-0-UI.r1 NIH_BMAP_FR0 Mus musculus cDNA clone
            UI-M-FR0-cbe-1-08-0-UI 5', mRNA sequence.
ACCESSION   BU615139
VERSION   BU615139.1 GI:23281354
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 706)
  AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
  TITLE   NIH-MGC http://mgc.nci.nih.gov/.
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr, Jim Lin, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            The following repetitive elements were found in this cDNA
            sequence: 218-261, >(GAAAA)n#simple_repeat (matched compliment)
            Seq primer: pYX-5.

FEATURES   Location/Qualifiers
     source   1..706
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone_lib="NIH_BMAP_FR0"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (TI phage resistant)"
                     /clone="UI-M-FR0-cbe-1-08-0-UI"
                     /note="Organ: Brain; Vector: pYX- Asc; Site 1: EcoR I;
                     Site 2: Not I; The library was constructed according
                     to Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."

ORIGIN

Query Match   54.3%;   Score 440.8;   DB 13;   Length 706;
Best Local Similarity 81.2%;   Pred. No. 2.1e-115;
Matches 575;   Conservative 0;   Mismatches 94;   Indels 39;   Gaps 4;

Qy   23 CGGCAAACTATAAACTGATTCTCATCAGGAAAACTGCACATTATCTCCCATCACTTCAAAG   82
             |||||||||| |||||||| |||| ||||||||| ||| |||||||||||| ||||||||
Db   6 CGTNGGCTATAAACTGCTTCTCATCACAGAAAACT--ACAATAGCTCACCATCACTTCAAAG   63

Qy   83 GTCTCCGTCAGGCACAGAGGTGACGCGCCAGGAGGAGATGATTTAAAGGTGAAAATGACAAGGTTTCC 142
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 GTCTCGTCAGACACGAGGGTGACGCCCAGGGGGGATGATTTAAAGGTGAACATGACAAGGTTTCC 123

Qy   143 ACCCCTCAAACCCTTGGCTCCTTTTTCTGACAATACCAGTCTGAATCAAACCCGATGTCTTTTT 202
             ||||||||||||||||||||||||||||||||||||| ||||||||||||||| ||||| ||
Db   124 [...] 202

```
Db    124  ACCCCTCAAACCTTGCTTCCTTCTTTCTGACAATACAG--TGAACGAGCCCAAGTC---TT 178
Qy    203  TTTTACTGTGGAAATAGGATCGGGAAGAGACTAAC------------|||||||||||||| 236
Db    179  CACACTGTGGAAACCGGACTTGGGAAGAAGAAAAGCTTGCCTTTTTTTTTCTTTTTTCTTT 238

Qy    237  ------ATTTTTTTTTTTTAATCCTGATAAAGAAAGAGATTGTTGGGAAGCTTCTTTGAAAAAAA 290
Db    239  TCTTTTTTTTTTTTTTTCAATTCTTAATAAATAGGAAT TGGGAAGCATTCTCTTTGAAAAAT 298

Qy    291  TTTTAAAATTGTGGGACAGAGATGCGGATTTTAGAAAAAGTTTAGATCTTTCCAATGAACACTAAT 350
Db    299  CTCAGAACGTGGCACAGAATGGATTTTAAAAAAGTGTTAGCTCTTTCCAATGAGCACTAGG 358

Qy    351  AGAGTAACTCTGCTTCTTGGCTGGATTTTTTCAGAGAATGGAATGGTTCTCTGCGATGTCCTG 410
Db    359  AGGGTTCCCGCTCTTGGCTGGATTTTTTCAGAGAATGGGAATGGTTCTCTGCGATGTCCTG 418

Qy    411  GGTCCGTATTTGTGGATAAGTGCTTGTGCAAATGCTACTCTGCCATGGATCCCTTCAGACA 470
Db    419  GGCCCTGTACTTGTGGAATAAGTGCTTGTGCCAATGGTCACTCCAACA 478

Qy    471  CACTTTCCAGCAGCATCACCTGCACAGACCAGAGGACGCGTGTGAAGTGAATGAGTAGCAGC 530
Db    479  CACCTTCCAGCAGCATCACCTGCACAGGACCCTGTGAAGTGAATGAGTGATCGCGGC 538

Qy    531  ACACCGATGTGTTGTAACAAGAATCGCATTGAGGAGCGGCGGTCACAAACAGTAAAGTGTTCCTG 590
Db    539  CCACAGGTGTGTTGTAACAGGAACCGATCGCAGAAAACAGTAAAGTGTTCCTG 598

Qy    591  TCTACCCTGGAAAAGTGGCTGGAACAACAAGAAACCGGCGGTTCTTCTTCTGGCGTCGAT 650
Db    599  TTTACCCTGGGGAAAGTGGCTGGGACAACAAGAAACCGGACTTCCTGTGTGGATGCCTCCAT 658

Qy    651  AGTGATTGGGGAAATGGTGGGTGTGTGAGGAGGCCTTGCCTAGGAGGAGGAGA 698
Db    659  AGTAATTGGGAAATGGTGGGTGTGTGAGGATGGAGCGGCCCTCGCCTANAAGGAGA 706
```

```
ORIGIN

Query Match      53.5%;  Score 434.8;  DB 9;  Length 975;
Best Local Similarity 95.6%;  Pred. No. 1.2e-113;
Matches 436; Conservative 9; Mismatches 11; Indels 0; Gaps 0;

Qy    357  CTCTGCTCTTGGCTGGGATTTTTCAGAGAATGGGAATGGTTCTCTTGCGATGTCTTGGGGTCCT 416
Db    411  CTTTCCTGTGTGGAATGTTCTCTTTAGAGAATGGGAATGGTTCTCTGCGGATGTCCTGGGGTCCT 470

Qy    417  GTATTTGTGGATAAGTGCTTGTTGTGCAAATGCTACTCTGCCATGGATCCCTTCAGACACACTTT 476
Db    471  GTATTTGTGGATAAGTGCTTGTTGGCAAATGCTACTCTGCCATGGATCCCTTCAGACACACTTT 530

Qy    477  CCAGCAGCATCACCTGCACAGACCAGAGGACGCGTGTGAAGTGAATGAGTAGCAGCACACCG 536
Db    531  CCAGCAGCATCACCTGCACAGACCAGAGGACGCGTGTGAAGTGAATGAGTAGCAGCACACCG 590

Qy    537  ATGTGTTGTAACAAGAATCGCATTGAGGAGCGGCGGTCACAAACAGTAAAGTGTTCCTGTCTACC 596
Db    591  ATGTTGTAACAGGAACCGATCGCAGAAAACAGTAAAGTGTTCCTGTCTACC 650

Qy    597  TGGGAAAAGTGGCTGGAACAACAAGAAACCGGCGGTTCTTCTTCTGGCGTCGATGAT 656
Db    651  TGGGAAAGTGGCTGGGACAACAAGAAACCGGACTTCCTGTGTGGATGCCTCCATAGTGAT 710

Qy    657  TGGGAAAATGGTGGGTGTGTGAGGAGGAGGCCTTGCCTAGGAGGAGGAGAGTGTAAGACACTCCC 716
Db    711  TGGGAAAATGGKGKGGKGTGAGGATGGAGCCTTKCCTAKAAGGAGGAGGAGAGTGTAAGACACTCCC 770

Qy    717  TGACAATTCTGGATGGATGTGGCGCAACAGGCAACAGAAAAATTAAGACCACGAGAATTCACCC 776
Db    771  TGACAATTCTGGATGGATGGCGRTGTGGCGCAACAGGCAACAGAAAAATTAAGACCACGAGAATTCACCC 830

Qy    777  AAGAACCTAACAGAAGCATTTTGTGTAGTAAAGGAA 812
Db    831  AAGAACCTAACAGAAGCATTTGTGGKXAGTAAAGGAA 866
```

```
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 948    Std Error: 0.00
Seq primer: M13RPI
High quality sequence stop: 360.
```

FEATURES   Location/Qualifiers
```
source    1..468
          /organism="Homo sapiens"
          /mol_type="mRNA"
          /db_xref="GDB:424929"
          /db_xref="taxon:9606"
          /clone="IMAGE:51993"
          /sex="female"
          /dev_stage="73 days post natal"
          /lab_host="DH10B (ampicillin resistant)"
          /clone_lib="Soares infant brain 1NIB"
          /note="Organ: whole brain; Vector: Lafmid BA; Site 1: Not
          I; Site 2: Hind III; 1st strand cDNA was primed with a Not
          I - oligo(dT) primer [5'
          AACTGGAAGAATTCGCGGCCGCAGGAArrTTTTTTTTTTTTT 3'];
          double-stranded cDNA was ligated to Hind III adaptors
          (Pharmacia), digested with Not I and directionally cloned
          into the Not I and Hind III sites of the Lafmid BA vector.
          Library went through one round of normalization. Library
          constructed by Bento Soares and M.Fatima Bonaldo."
```

ORIGIN

```
Query Match    50.4%;    Score 409;    DB 14;    Length 468;
Best Local Similarity   98.6%;    Pred. No. 2.7e-106;
Matches  412;  Conservative  0;  Mismatches  6;  Indels  0;  Gaps  0;

Qy  395  TCTCTCGGATGTGCCTGGGTCCTGTATTTGTGGATAAGTGCTTGTGTGCCAATGCTACTCTGCC  454
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  TCTCTCGGATGTCCTGGGTCCTGTATTTGTGGATAAGTGCTTGTGTGTCAATGCTACTCTGCC   60

Qy  455  ATGGAtCCCTTCAGCACACTTTCCACAGCATCACCTGCACAGACCAGAGAGAGGACGT  514
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  ATGGAtCCCTTCAGCACACTTTCCACAGCATCACCTGCACAGACCAGAGAGAGGACGT  120

Qy  515  GTGAAGTGATAGCAGCAGCACCACCGATGTTGTAACAAGAATGTCGCATTGAGGACGGGTCACAA  574
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  GTGAAGTGATAGCAGCAGCACCACCGATGTTGTAACAAGAATGTCGCATTGAGGACGGGTCACAA  180

Qy  575  CAGTAGAAGTGTTCCTGTGTCTACCTGGGAAAAGTGGCTGGAACAACAGAACCGGCCTTCTT  634
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  CAGTAAAGTGTTCCTGTGTCTACCTGGGAAAAGTGGCTGGAACAGAGAACCGGCCTTCTT  240

Qy  635  GCGGTCGAtGCCTCCCATAGTGATTGGGAAAATGTGGGTGTGAGATGGAGCCTTGCCTAGGAAG  694
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  GCCGTCGAtGCCTCCCATAGTGATTGGGAAAATGGGTGGTGTGAGAtGGAGCCTTGCCTAGGAAG  300

Qy  695  GAGAGAAGTGTAAGACACTCCCTGACAAATTCTTGGAtGGAtGGGCGCAACAGCCACACAAAA  754
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GAGAGAAGTGTAAGACACTCCCTGACAAATTCTTGGAtGGAtGGGCGCAACAGCCACACAAAA  360

Qy  755  TTAAGACCCACGAGAATTCACCCAAGAACCTACAGACACATTTGTGGTGAGTAAAAGGAA  812
              ||||||||||||||||||||||||||||||||||||  |||||  ||||||||||||
Db  361  TTAAGACCACGAGAATTCACCCAAGAACCTACAGACAAGGCTTTGGTTGTTGTAGTAAAAGGGA  418
```

RESULT 13
BB653506
```
LOCUS       BB653506       692 bp    mRNA    linear    EST 26-OCT-2001
DEFINITION  BB653506 RIKEN full-length enriched, adult male hippocampus Mus
            musculus cDNA clone C630045F22 5', mRNA sequence.
ACCESSION   BB653506
VERSION     BB653506.1  GI:16487345
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1  (bases 1 to 692)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiranoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
  COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp;
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,
            Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
            e mouse tissues.
```

FEATURES   Location/Qualifiers
```
source      1..692
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="C630045F22"
            /sex="male"
            /tissue_type="hippocampus"
            /dev_stage="adult"
            /lab_host="SOLR"
            /clone_lib="RIKEN full-length enriched, adult male
            hippocampus"
            /note="Site 1: XhoI; Site 2: BamHI; cDNA library was
            prepared and sequenced in Mouse Genome Encyclopedia
            Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in
            RIKEN. Division of Experimental Animal Research in Riken
            contributed to prepare mouse tissues. 1st strand cDNA was
            primed with a primer [5'
            GAGAGAGAGAGAGAGACTCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse
            transcriptase and subsequently enriched for full-length by
            cap-trapper. Second strand cDNA was prepared with the
            primer adapter of sequence [5'
            GAGAGAGAGAGAGAGA 3']"
```

ORIGIN

```
Query Match    47.9%;    Score 389.2;    DB 10;    Length 692;
```

RESULT 14
AK049880

LOCUS   AK049880   2412 bp   mRNA   linear   HTC 20-SEP-2003
DEFINITION   Mus musculus adult male hippocampus cDNA, RIKEN full-length
               enriched library, clone:C630007B19 product:unclassifiable, full
               insert sequence.
ACCESSION   AK049880
VERSION   AK049880.1 GI:26093679
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99272253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,

REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
               Konno,H., Akiyama,J., Nishi,K., Kitsuna,T., Tashiro,H., Itoh,M.,
               Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
               Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,K., Kashiwagi,K.,
               Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
               Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
               Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
               sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4 (bases 1 to 2412)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
               FANTOM Consortium.
  TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)

REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
               Group Phase I & II Team.
  TITLE   Analysis of the mouse transcriptome based on functional annotation
               of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)

REFERENCE   6 (bases 1 to 2412)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
               Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
               Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
               Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
               Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
               Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
               Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
               Okazaki,Y., Saito,R., Saito,H., Sakai,C., Sakai,K., Sakazume,N.,
               Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
               Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
               Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
               Muramatsu,M. and Hayashizaki,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
               Physical and Chemical Research (RIKEN), Laboratory for Genome
               Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
               RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
               Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
               Tel:81-45-503-9222,
               Fax:81-45-503-9216)
COMMENT   cDNA library was prepared and sequenced in Mouse Genome
               Encyclopedia Project of Genome Exploration Research Group in Riken
               Genomic Sciences Center and Genome Science Laboratory in RIKEN.
               Division of Experimental Animal Research in Riken contributed to
               prepare mouse tissues.
               Please visit our web site for further details.
               URL:http://genome.gsc.riken.go.jp/
               URL:http://fantom.gsc.riken.go.jp/.
FEATURES   Location/Qualifiers
     source   1..2412
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:C630007B19"
                     /db_xref="MGI:2417405"
                     /db_xref="taxon:10090"
                     /sex="male"
                     /clone="C630007B19"
                     /tissue_type="hippocampus"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     misc_feature   1..2412

TITLE
JOURNAL
  MEDLINE
   PUBMED
REFERENCE
  AUTHORS   Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
               Normalization and subtraction of cap-trapper-selected cDNAs to
               prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159

Best Local Similarity   79.3%;
Matches 526; Conservative 0; Pred. No. 1.4e-100;
Mismatches 99; Indels 38; Gaps   4;

Qy   2 ACTGGAGTGCGGGATGGTGGTCCACCGGCAACATAAACATAAACTGATTCTCATCAGGAAACTGCACAT 61
           ||||||||| |||||||| ||||||||||||| ||| |||| |||||||||||||||||||||
Db   31 ACTGGAGTAGGGATGGCATGGTCGTTGTCCGTTGGCTTAATCGTTCTCATCAGGAAACT--ACAT 88

Qy   62 TATCTCCCCATCACTTCAAAGGTTCTCGTCGTCAGCGACGACGAGGTGACGCCCAGGAGAGTGAGTTTAA 121
           ||||||| ||||||||||||||||||||| ||||| |||||||||||||||||| |||||||||| ||||
Db   89 TAGCTCACCAATCACTTCAAAAGGTCTCGTCGTCAGCAGAGGTGACGCCCAGGGGATGATTTAAA 148

Qy   122 GGTGAAAATAGACAAGGTTCCACCCCTCAAAACCTTGGCTCCTTTTCTGACAATACAGTCT 181
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   149 GGTGAACATGACAAGGTTTCCACCCCTCAAAACCTTGGCTTCCTTTTCTGACAATACAG--T 206

Qy   182 GAATGAACCCGATGTGTCTTTTTTTTTGTGGAAATAGGATCGGGAAGAGAGAGTAAC----- 236
           ||||||||| ||| ||||||||||||||||||||||| ||||| |||||||
Db   207 GAACGAGCCCAACGTC---TTCACAACTGTGGGAAACGGGACTGGGAGAGAGAAAAGCTTGCC 263

Qy   237 ------------------AGTTTTTTTTTAATCCTGATAAAGGAAGAAGATTGTTG 270
                                 ||||||||||||| ||| ||||| ||| |||
Db   264 TTTTTTTTTCTTTCTTTTTTTTTTTCAATTCTTAATTAAACGAATTGCTGTGG 323

Qy   271 GGAAGCTCTTTGAAAAAAATCTCAGAACTGTGGCACAGATGGATTTTAAAAAGTGTTAGA 330
           ||||||||||||| |||||||||||||||||||||||||||||| |||||||||||||||
Db   264 TTTTTTTCTTTCTTTCTTTTTTTTTCAATTCTTAATTAAACGAATTGCTGTGG 323

Qy   324 AAGCAATTCTTCTTTGAAAAAATCTCAGAACTGTGGCACAGATGGATTTTAAAAAGTGTTAGC 383
           ||||||||||| |||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   271 GGAAGCTCTTTGAAAAAAATTTAAAATGTGGGCACAGATGGATTTTAAAAAGTGTTAGA 330

Qy   331 TCTTTTCCAATCAACACTAATAGAGTACTCTGCTTCTTTGGCTCTTGGCTTTTCAGAGAAATGGCA 390
           |||||||||||||||||||||| |||||||||||||||||||||| ||| |||||||||||||
Db   324 AAGCAATTCTTCTTTGAAAAAATCTCAGAACTGTGGCACAGATGGATTTTAAAAAGTGTTAGC 383

Qy   384 TCTTTTCCAATCAACGACGCACTAGGAGGGGGTTCCCTCTTGGCTGGATTTTTCAGAGAATGGCA 443
           |||||||||||||||||| ||||| |||||||||||||||||||||||||||||||||||||
Db   331 TCTTTTCCAATCAACACTAATAGAGTACTCTGCTTCTTTGGCTCTTGGCTTTTCAGAGAAATGGCA 390

Qy   391 ATGGTCTCTGCCAATGGTCCTCTGGGCTCCTGTATATTGTGGATAAGTAGTGCTTGTGTCAATGCTACTC 450
           |||||||| ||||||||| ||||||||||||| |||||||||| ||||||||| ||||||||||||||
Db   384 TCTTTTCCAATCAACGACGCACTAGGAGGGGGTTCCCTCTTGGCTGGATTTTTCAGAGAATGGCA 443

Qy   444 ATGGTCTCTGGAATGTCCTGGGCCCTTGTGGATTAAGTGTCTTGTTGCCGATGTGCTC 503
           |||||||||||||| ||| |||| ||||| |||| ||| |||||||||| ||| |||
Db   391 ATGGTCTCTGCCAATGGTCCTCTGGGCTCCTGTATATTGTGGATAAGTAGTGCTTGTGTCAATGCTACTC 450

Qy   451 TGCCATGGATCCCTCTCAGCACACTTTTCCAGCAGCAATACCACGTGGACCAGGAGGAGG 510
           |||||||||||||||||||| |||||||||||||||||||||||||||||||||| |||
Db   444 ATGGTCTCTGGAATGTCCTGGGCCCTTGTGGATTAAGTGTCTTGTTGCCGATGTGCTC 503

Qy   504 TGCCATGGGTCACTCCAACAACACCTTCCAGCAGCAATGACCAGGAAGGAGGG 563
           |||||||| |||||||||||||||||| |||||| ||||||||||||| |||||||
Db   451 TGCCATGGATCCCTCTCAGCACACTTTTCCAGCAGCAATACCACGTGGACCAGGAGGAGG 510

Qy   511 ACGTGTGAAGTGAATAGCGACACACCGATGATTGTAACGAAGAAATCGCATTGAGGAGGGTCA 570
           |||||||||| |||||||| ||| |||||||||| ||||||| ||||||||||||| |||||
Db   504 TGCCATGGGTCACTCCAACAACACCTTCCAGCAGCAATGACCAGGAAGGAGGG 563

Qy   564 ACTTGTGAAGTGAATCGCGGGCCCCACAGTGTTGTTGTAACAAGGAACCGGATCGAGGGGCGGGTCA 623
           |||||||||||| ||| |||||||||||| |||||||||||||| |||||||||
Db   511 ACGTGTGAAGTGAATAGCGACACACCGATGATTGTAACGAAGAAATCGCATTGAGGAGGGTCA 570

Qy   571 CAAACAGTAAAGTGTTCCTGCTTGTCTTACCTGGAAAAAGTGGCTGGGAACAACAAGGAAACCGGGCCT 630
           ||||||||||||| ||||||||||||||| ||||||||||||||||| |||||||||||| |||||||
Db   564 ACTTGTGAAGTGAATCGCGGGCCCCACAGTGTTGTTGTAACAAGGAACCGGATCGAGGGGCGGGTCA 623

Qy   624 CAAACAGTGAAGTGTTCCTGTTTTACCTGNGAACAGAAAGTGGCTGGGGAACAGAACCGACCTT 683
           |||||||| ||||||||||||||||| ||| ||||||||||||||||||||||| |||| |||
Db   571 CAAACAGTAAAGTGTTCCTGCTTGTCTTACCTGGAAAAAGTGGCTGGGAACAACAAGGAAACCGGGCCT 630

Qy   631 TCT 633
           |||
Db   684 CCT 686

Qy   CAAACAGTGAAGTGTTCCTGTTTTACCTGNGAACAGAAAGTGGCTGGGGAACAGAACCGACCTT 683
Db   624

```
FEATURES             Location/Qualifiers
     source          1..845
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="Compton Line 151"
```

```
COMMENT         Contact: Simon Hubbard
                Department of Biomolecular Sciences
                University of Manchester Institute of Science and Technology
                (UMIST)
                PO Box 88, Manchester, M60 1QD, UK
                Tel: 0161200930
                Fax: 01612360409
                Email: Simon.Hubbard@umist.ac.uk.
MEDLINE    22233534
PUBMED     12445392
TITLE      A Comprehensive Collection of Chicken cDNAs
JOURNAL    Curr. Biol. 12 (22), 1965-1969 (2002)
AUTHORS    Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
           Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
REFERENCE  1  (bases 1 to 845)
           Phasianinae; Gallus.
ORGANISM   Gallus gallus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
SOURCE     Gallus gallus (chicken)
KEYWORDS   EST.
VERSION    BU275101.1  GI:25546051
ACCESSION  BU275101
           sequence.
DEFINITION 603532890F1 CSEQCHN53 Gallus gallus cDNA clone ChEST489d23 5', mRNA
LOCUS      BU275101             845 bp    mRNA    linear    EST 26-NOV-2002
BU275101
RESULT 15
```

```
Db   731 AGAAGCAATTGTTATATAAAATAGGA 755
         ||||||||||||||||||||||||||
Qy   788 AGAAGCAATTGTGGTAGTAAAGGAA 812

Db   671 GATGGAAGTGTGTCTACAGGCAACAAGATTAAGACTACACGAATTCACCCAAGAACCTAAC 730
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   728 GATGGAAGTGTGCGCAACACAGGCAACAAAAATTAAGACACCAGAGAATTCACCCAAGAACCTAAC 787

Db   611 GGTGTGAGATGGAGCCCCTGCCTAGAGGAGGAGAATGTAAGACACTCCCCTGACAATTCTG 670
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   668 GGTGTGGATGGAGTGGAGCCTTGCCTAGAGAGAGAGAGAGAGAATGTAAGACACTCCCCTGACAATTCTG 727

Db   551 CTGGGACAACAAGAAACCGGACCTTCCTGTGTGGAGCGTCCCATAGTAATTGGGAAATGGT 610
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   608 CTGGAACAACAAGAAAACGGGCCTTCTTCTTCTTGCCTGATGCCTCCCATAGTGATTGGGAAATGGT 667

Db   491 AGAAACCGCATCGAGTGCACATCAGTGAAGTGTTCACTGGAAGAAAGTGGACA 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   548 AGAAATGGCATTGAGTGACGAGCGGGTCACACAAACATAAAGTGTTCCTGTTACTGTTACTGGGAAAGTGG 607

Db   431 ACTTGCACCGGCCACAGAAGAGGACCTGTGAAGTGATCGGCGCGCCCACAGCTGTGTTGACA 490
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   488 ACCTGCACAGACCAGGAGGAGGACCGTGTGAAGTGATTAGCAGCACACCGGATGGTTGTAACA 547

Db   371 TAAGTGCTTGTTGCGATGCTGCTCTGCCATGGGGTCACTCCAACACACTTCCAACGATC 430
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   428 TAAGTGCTTGTTGTGTGCAAATGCTACTCTCTGCCATTCTGCCAATGTCCCCTTGGGCGCCCTTGACTTGTGGA 487

Db   311 GATGTTCTCTTTAGAGAAT GGCAATGGTCTTCTGCAATGTCCCTGGGGCCCTGTACTTGTGTGGA 370
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   368 GCTGGATTTTTCAGAAGATGGCAATGGTCTCTGCGAGTGTCCTCGGTATTGTTTGTGGA 427
```

```
ORIGIN
     /db_xref="taxon:9031"
     /clone="ChEST489d23"
     /sex="female"
     /tissue_type="cerebrum"
     /dev_stage="adult"
     /lab_host="DH10B"
     /clone_lib="CSEQCHN53"
     /note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
     EcoRI; Site 2: NotI; This normalized library was
     constructed from 1 million independent clones. cDNA
     synthesis was initiated using an oligo(dT) primer, using
     methylated C in the first strand synthesis reaction.
     Following this first strand reaction, double-stranded cDNA
     was blunted, ligated to NotI adapters, digested with
     EcoRI, size-selected, and cloned into the NotI and EcoRI
     compatible sites of a custom modified MCS of the
     pBluescript (KS+) vector. The library was normalized in 2
     rounds using conditions adapted from Soares et al., PNAS
     (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
     (1996):791, except that a significantly longer
     reannealing hybridization was used."
```

```
Query Match       44.6%;   Score 361.8;  DB 13;   Length 845;
Best Local Similarity  77.1%;  Pred. No. 1.1e-92;
Matches  565;  Conservative   0;  Mismatches  112;  Indels  56;  Gaps  8;

Qy    47 CAGGGAAACTGCACATTATCTCCCATCACTCACTTCAAAGGTCTTCGGTCAGGCAG-AGGTGACGC 105
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    83 CTGGGAAGCTTGCAAAATTGTCTCGAGATCGCTTGTCAGATGGGAATCGCTTTATCAAGCAGAAGGTGATGC 142

Qy   106 CAGGAGAGATGATTTAAAGGTGAAAATGACAAAGGTTTCACCCCTCAAAACCTTGGCTCCTTT 165
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   143 TAGGAGATGATTTAAAGGTGAAAAATGAGAGAAGGTTCCACCCCTTCAAAACCTTCGGCCACTTT 202

Qy   166 TCTGACAATACAGTCTGAATGAACCCGATGCTCTTTTTTTTTACTGTGGAAATTAGGGATCGG 225
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   203 TCTGACGGCAGTGCCTGACAG----CTCAGATGTCGTCTTTACTCTGTGGGAACCAGGATTTG 258

Qy   226 AAGAGAGTAACATTTTTTTTTTTT-----TAATCCTGATAAAGAGAAGATTGTTGGGGAAGCTCTTT 280
         |||||||||||||||||||           |||||||||||||||||||||
Db   259 TAACAAGAGAGATTTTTTTTTTCCTCGTAATCCTGAATCCTGATAAAAGAAGATTATGGGGAAGCTATT 318

Qy   281 TGAAAAAAAAAATTTTAAATTGTGGGCACAGATGGATTTTAAAAGTTGGGATCTTTCAAT 340
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   319 -----AAAAATCCCTTATTCTGTGGTCGCAGATGGATTTTTAAAAAGTGTTAGAGACCTTTCAAT 373

Qy   341 GAACACTAATAGAGTACTCTGCTCTCTGGCTCTGGGATTTTCGAGAGAATGGCAAATGGTCTCTG 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   374 GAACACTAATAGAGTGCTCTGCTCTTGGCTGCGGGAATTACTCAGAGAAATGGGCAGTGGTCTCTG 433

Qy   401 CGATGTCCTCGGGTCCTGTATTTGTGATTAAGGTTGCATGTGCAATGTCTACTCTGCCATGGAT 460
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   434 CAATGTCCTGGGTCCTGTATTCGTGGATAAGTGCCTGTGCAATGCTTGTCTGCAGGGAT 493

Qy   461 CCCTTCAGCA--CACTTTCCAGCAGCATCACTGCACACGAACCAGAA------------- 504
         |||||||||||  |||||||||||||||||||||||||||||||
Db   494 CCCTCCATCAACCACTTTCCAGCAGCATCACCTGCACAGACCAGGAAAAGACTCTTTCTGT 553

Qy   505 ---------------------GGAGGGACGTGTGGAAGTGAAGTGAGTAAGCAGCACACCGGATGTT 541
                              ||||
Db   554 GAATCAGAGATTTCCCAGTGCGGGAAGGCGGTGCAGGACGACGGTGTGAGAGTTGAAGTTATGCAGCACACAGATGCT 613

Qy   542 GTAACAAGAANTCGCAATTGAGGAGCCGGTCACAAAACAGTAAAGTGTTCCTGTCTACCTGGAA 601
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   614 GTAATAAGAANTCGAATTGAGGAAAGATCACAAACAGTAAAAATGCTCGTCTACCTGGGA 673

Qy   602 AAGTGGCTGGGAACAACAAGAAAACCGGCCTTCTTGCGTCGATGCCTCCATAGTGATTGATTTGGGA 661
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   674 AAGTGGCTGGGACTACACGAAAACAGACACTTCCCTGTTGTGANGCTTCAAATAGTGATGGGGA 733

Qy   662 AATGGTGGGTGT-GAGATGGAGCCCTTGACCTGAGAAGGAGAGAGAAGAAATGTAAGA-CACTCCCTGA 719
         || |||| |||  |||||||||||||||||||||||||||||||||||  |||||||||
Db
```

```
Db   734 AATGGTGGTGTGTGGAAATGGGAGCCTGCCTAGAAGGGAGAGGAAATGTACGACCGTTGCCTGA 793
Qy   720 CAATTCTGGATGG 732
             || ||||||||||
Db   794 TAAAATCTGGATGG 806
```

Search completed: April 9, 2004, 15:45:13
Job time : 2566 secs